The assessment of fatigue
A practical guide for clinicians and researchers

A.J. Dittner a, S.C. Wessely b, R.G. Brown a,*

a Department of Psychology, Institute of Psychiatry, King’s College London, London, UK
b Academic Department of Psychological Medicine, Guy’s King’s and St. Thomas’ School of Medicine, London, UK

Abstract

Objectives: Fatigue is a common feature of physical and neurological disease as well as psychiatric disorders, often reported amongst patients’ most severe and distressing symptoms. A large number of scales have been developed attempting to measure the nature, severity and impact of fatigue in a range of clinical populations. The aim of the present review is to guide the clinician and researcher in choosing a scale to suit their needs. Methods: Database searches of Medline, PsycINFO and EMBASE were undertaken to find published scales. Results: Details of 30 scales are reported. These vary greatly in how widely they have been used and how well they have been evaluated. The present review describes the scales and their properties and provides illustrations of their use in published studies. Conclusions: Recommendations are made for the selection of a scale and for the development and validation of new and existing scales.

Keywords: Measurement; Instrument; Scale; Psychometric; Severity; Impact

Introduction

Although often identified as a sign or symptom of a disease state or side effect or treatment, fatigue is essentially a subjective experience. It has largely defied efforts to conceptualise or define it in a way that separates it from normal experiences such as tiredness or sleepiness. Emphasis is usually given to the degree and persistence of such experiences in the absence of any excessive expenditure of energy or effort as cause. Thus, fatigue is typically defined as extreme and persistent tiredness, weakness or exhaustion—mental, physical or both. Fatigue is common in the general population [1,2] and is the defining feature of chronic fatigue syndrome (CFS). However, it is also an important feature of a wide range of other conditions including physical disease such as cancer, neurological disease such as multiple sclerosis (MS) and Parkinson’s disease and psychiatric disorders such as depression. In these and other conditions, fatigue can be a major source of disablement and is often reported by patients as being amongst their most severe and distressing symptoms [3–7]. Despite this, fatigue has typically been ignored in the assessment of symptom severity or outcome in many of the diseases in which it is found. Consequently, we know little about the phenomenology of fatigue in these conditions, quite apart from their epidemiology and aetiology. Finally, fatigue is often neglected as a target for treatment, perhaps because it typically appears unrelated to the severity of the central disease process.

Progress in research and improved management depends on having reliable and valid methods of assessment that reflect the problems reported by patients. With the growing recognition of fatigue as a major clinical problem in many conditions, there has been a proliferation of measures of fatigue, often referred to by synonyms or abbreviations shared with other scales. Although all purport to assess fatigue, being self-report scales, the information derived depends on the questions being asked. These will be based on the scale developer’s own conceptualisation of fatigue and will in turn be answered by the respondent based on his or her own interpretation. This means that different scales may be measuring fundamentally different aspects of the fatigue experience or even potentially distinct constructs. In addition, where an instrument has been developed specifi-
cally to measure fatigue in one clinical condition, its use in other patient groups may not be justified if the fatigue experience differs from group to group.

A researcher or clinician wishing to measure fatigue in their patients needs to ensure that the instrument chosen measures the right aspect of fatigue for their purposes, in a way that meets the requirements of their study and does so both reliably and validly. However, choice of the most appropriate measure is far from straightforward. The purpose of the present review is to describe the range of instruments available and to provide guidance on choosing a scale for a specific use. It does not seek to compare scales directly although published studies that have sought to do so will be discussed.

Procedure

The scales included in this article are the result of a bibliographic search of English language publications indexed in Medline (1966 to March 2003), EMBASE (1980 to March 2003) and PsyclINFO (1974 to March 2003). Searches were based on the main Medical Library Subject Heading (MESH) term “fatigue” (synonym “lassitude,” previously “tiredness”). The scope of this term is defined as “the state of weariness following a period of exertion, mental or physical characterised by a decreased capacity for work and reduced efficiency to respond to stimuli.” It is distinguished from “muscle fatigue” defined in MESH as “a state arrived at through prolonged and strong contraction of a muscle.” In addition to the search on the main term “fatigue,” a parallel search was also made on the conceptually related term “asthenia,” defined in MESH as a “clinical sign or symptom manifested as debility or lack or loss of strength and energy.” Other related constructs such as “tiredness” and “anergia” are not considered as distinct signs and symptoms in the MESH classification system. The scope of the review excluded scales designed to assess sleepiness or somnolence.

For the Medline and EMBASE searches, the MESH qualifier “/Diagnosis” was used. This covers all aspects of diagnosis, including examination, differential diagnosis and prognosis.Qualifiers are not available with PsyclINFO. Therefore, the MESH terms “fatigue” and “asthenia” were combined with a keyword search of “instrument,” “assessment,” “scale” or “measurement.” This search was augmented by reviewing article reference lists and performing citation searches using ISI Web of Science. Scales cited only in abstracts or as reports of meetings were not included.

Details of all scales identified are presented in Tables 1 and 2. These tables summarise each scale’s purpose and structure and evidence of its psychometric properties from the original source reference. Where available, published cutoff scores are provided for guidance, although their validity or utility in other clinical or research contexts should not be assumed. For the majority of scales, further details including illustrations of their published uses are provided in the accompanying text. Where psychometric properties were not explicitly tested in the primary reference, potential users may need to check for any subsequent information pertaining to reliability and validity. The order of presentation is alphabetical, commencing with unidimensional scales (Table 1) and then multidimensional scales (Table 2). Scales for which insufficient are data available at present are included in the tables but discussed only briefly in a final section on “Other scales.”

The scope of this review excludes instruments that include fatigue as one dimension of broader index of health outcome. These include generic instruments such as the Medical Outcomes Study Short Form-36 (SF-36) [8], the Nottingham Health Profile (NHP) [9] and the Profile of Mood States (POMS) [10] in addition to many diseasespecific general outcome scales. Such measures can provide a useful brief index of fatigue in the context of broader health outcome. However, the fatigue subscales or items should generally not be used in isolation without validation, although the POMS Fatigue subscale has been used independently in many studies [11].

The following comments and discussions should be read in conjunction with the details reported in Tables 1 and 2 together with the recommendations provided at the end of the review.

Unidimensional scales (Table 1)

The Brief Fatigue Inventory (BFI) [12]

The BFI was developed for screening and assessing clinical outcome in severely fatigued patients with cancer. The authors acknowledge that the scale is virtually interchangeable with other unidimensional fatigue severity scales such as the Functional Assessment of Cancer Therapy-Fatigue (FACT-F) (see below) but claim its use of language is simpler making it easier both to understand and to translate. The BFI has good psychometric properties although, at the time of writing, there is no information on test–retest reliability or its sensitivity to change. It has not been used in any subsequent studies and has not been formally validated in a noncancer population.

Number of citations: 23.
Examples of use: cancer [13].

Fatigue Severity Scale (FSS) [14]

This is one of the best known and most used fatigue scales. The name is, however, slightly misleading. The FSS principally measures the impact of fatigue on specific types of functioning rather than the intensity of fatigue-related symptoms [15].

The FSS has high internal consistency, has good test–retest reliability and is sensitive to change with time and after treatment. It also has good concurrent validity and is able to distinguish patients with different diagnoses (between systemic lupus erythematosus (SLE) and MS [14] and
between CFS, MS and primary depression [4]). In a comparison of the FSS and the Fatigue Questionnaire (FQ) (see later) in a sample of CFS patients, the FSS was found to be the more effective measure, probably owing to its specificity to the behavioural consequences of fatigue [15]. The scale’s psychometric properties have been confirmed in chronic hepatitis C [16] and immune-related polyneuropathies [17] although a study of fatigue in patients with brain injury [18] failed to support its internal consistency, suggesting that its suitability in all populations cannot be assumed.

Number of citations: 239.
Examples of use: MS [19,20], Parkinson’s disease [21], CFS [22,23], chronic hepatitis C [16], brain injury [18], sleep disorders [24], cancer [25,26] and amyotrophic lateral sclerosis [27].

**FACT-F subscale [28]**

The FACT-F has reasonable psychometric properties but is by definition a cancer scale and has not been validated in other populations. Although it has been validated independently from the full FACT scale and may be used in isolation, its diagnostic sensitivity and sensitivity to change have not yet been established.

Number of citations: 71.
Examples of use: cancer [29].

**Global Vigour and Affect (GVA) [30]**

This pair of measures, designed for research purposes, consists of eight 100-mm visual analogue scales, of which four are related to “vigour” and four to “affect.” The two subscales are scored individually so it is possible to derive a single score for global vigour (GV). In the initial validation study, GV was found sensitive to changes in mood and activation resulting from diurnal variations and jetlag. However, it was reported that subjects required time to practice and an explanation of the terms to complete the scale. This would make the measure unsuitable for postal surveys or any situation where unattended completion is required. Finally, the visual analogue scales make it laborious to score.

Number of citations: 56.
Examples of use: drug treatment effects [31] and sleep and circadian rhythms research [32].

**May and Kline Adjective Checklist [33]**

The scale consists of fatigue-related adjectives rated on a Likert scale. However, in the initial validation, scores on the checklist were found to correlate with scores on the Eysenck Personality Questionnaire (EPQ) [34], reflecting differences in personality rather than in fatigue. Whether this is a factor in only this questionnaire or in all measures of fatigue has not been assessed. Considering its length, it conveys relatively little information about the patient’s fatigue. In fact, many of the adjectives relate more to mood than to fatigue.

Number of citations: 7.
Examples of use: none to date.

**Pearson–Byars Fatigue Feeling Checklist [35]**

The Pearson–Byars Fatigue Feeling Checklist was developed to assess work-related fatigue in healthy adults and has been shown to discriminate between fatigued and non-fatigued airmen. It is not recommended for clinical or research use, owing to the lack of validity data in medical populations and its outdated language.

Number of citations: 12.
Examples of use: cancer [36,37] and pregnancy [38].

**Rhoten Fatigue Scale [39]**

The Rhoten Fatigue Scale has been used in a number of studies, mainly in patients with cancer, although because of its simplicity and generic language, it is likely to be useable in other conditions. As a single-item measure, it provides limited information about the patient’s fatigue, although it is useful as a quick screening measure.

Number of citations: 23.
Examples of use: cancer [40–42].

**Schedule of Fatigue and Anergia (SOFA) [43]**

First published in 1996 [44], the SOFA exists in two forms—the SOFA/CFS for the identification of patients with CFS in specialist clinics and the SOFA/GP, a modified version for the identification of prolonged fatigue syndromes in community and primary care settings. The scales differ in terms of their anchor points for severity and chronicity to optimise sensitivity to cases in the respective clinical settings for which they are intended. Both scales have good diagnostic validity, demonstrating their utility as screening instruments for patients with CFS and prolonged fatigue syndrome. However, the scales have not been developed or validated for use in other populations, and it remains to be seen whether they can be used to assess fatigue in other conditions.

Number of citations: 32.
Examples of use: CFS [45].

**Multidimensional scales (Table 2)**

**Chalder Fatigue Scale**

See FQ.

**Checklist Individual Strength (CIS) [46]**

The CIS was developed for use in hospital studies of CFS patients. As a multidimensional measure of severity and behavioural consequences of fatigue, CIS is divided into four subscales: Subjective experience of fatigue, Concentration, Motivation and Physical activity. The CIS has been well validated amongst CFS patients [46–49] and has been widely used in this population. It has good internal consistency and split-half reliability and discriminates amongst CFS, MS and healthy patients. Test–retest reliability has not been demonstrated, although the scale has been shown to be sensitive to change in fatigue levels over time and to drug treatment effects in randomised controlled trials.
The CIS has also been validated in the working population [50].

Number of citations: 57.

Examples of use: CFS [51, 52], MS [53] and working adults [50].

**Fatigue Assessment Instrument (FAI) [54]**

The FAI, sometimes referred to as the Fatigue Severity Inventory (FSI), is an expanded version of the unidimensional FSS (see above), with items added to assess additional aspects of fatigue. The scale was developed to permit the assessment of fatigue symptomatology across a range of medical conditions. It was therefore validated in a sample of outpatients at neurology and rheumatology clinics with a variety of diagnoses.

The FAI has four subscales: Fatigue severity, Situation specificity, Consequences of fatigue and Responsiveness to rest/sleep, with extra dimensions providing information on situational aspects of fatigue. The Fatigue severity subscale corresponds almost exactly to the FSS, sharing eight of the original nine items while including three new ones. Not surprisingly, its correlation with the FSS was found to be extremely high ($r = .98$) across 235 subjects in seven different disease groups. In general, the inventory has good psychometric qualities, although test–retest reliability is only moderate. Furthermore, closer examination of the factor structure indicates that the majority of the items loaded on to the first two factors and only Severity and Consequences subscales demonstrated concurrent validity based on other measures of fatigue and energy level. In its favour, however, the FAI is able to distinguish healthy subjects from patients and is notable for its ability to distinguish differences between patients with different diagnoses in some cases.

The FAI has been adapted [55] for use in Parkinson’s disease patients and has been named the FSI. Concurrent validity was demonstrated with the several other fatigue measures, although no other psychometric information was available at the time of writing.

Number of citations: 52.

Examples of use: chronic hepatitis C [56] and Parkinson’s disease [57].

---

Table 1
Unidimensional fatigue scales, characteristics and properties

<table>
<thead>
<tr>
<th>Scale name</th>
<th>BFI</th>
<th>CRFDS$^a$</th>
<th>DFIS</th>
<th>FSS, KFSS</th>
<th>FACT-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is assessed?</td>
<td>Severity</td>
<td>Impact</td>
<td>Impact</td>
<td>Impact and functional outcomes related to fatigue</td>
<td>Severity and impact</td>
</tr>
<tr>
<td>Number of scale items</td>
<td>9</td>
<td>20</td>
<td>8</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Scale type</td>
<td>11-point Likert</td>
<td>11-point Likert</td>
<td>5-point Likert</td>
<td>7-point Likert</td>
<td>5-point Likert</td>
</tr>
<tr>
<td>Number of subscales or factors</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Target population</td>
<td>Cancer inpatients, outpatients and community-dwelling adults without cancer diagnosis (595)</td>
<td>Cancer patients (221)</td>
<td>General medical individuals with flu-like illness (93)</td>
<td>Chronic medical MS and SLE patients (54)</td>
<td>Cancer patients receiving cancer treatment (49)</td>
</tr>
<tr>
<td>Standardisation sample(s) (n)</td>
<td>Cancer inpatients, outpatients and community-dwelling adults without cancer diagnosis (595)</td>
<td>Cancer patients (221)</td>
<td>General medical individuals with flu-like illness (93)</td>
<td>Chronic medical MS and SLE patients (54)</td>
<td>Cancer patients receiving cancer treatment (49)</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>0.96</td>
<td>0.97</td>
<td>0.91</td>
<td>0.88</td>
<td>0.93</td>
</tr>
<tr>
<td>Test–retest reliability</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.84</td>
<td>0.90</td>
</tr>
<tr>
<td>Concurrent validity</td>
<td>Associated with POMS-F and FACT-F</td>
<td>–</td>
<td>Negatively associated with health, sleep quality and activity; positively associated with illness symptoms, rating of fatigue and number of hours work missed</td>
<td>Fatigue rated on visual analogue scale</td>
<td>POMS-vigour, POMS-fatigue and prolonged fatigue syndrome</td>
</tr>
<tr>
<td>Discriminative validity</td>
<td>Discriminated between patients based on haemoglobin levels, subjectively rated fatigue and performance status</td>
<td>–</td>
<td>–</td>
<td>Distinguished patients with MS or SLE from healthy subjects</td>
<td>–</td>
</tr>
<tr>
<td>Cutoff score</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3/4</td>
<td>–</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>–</td>
</tr>
</tbody>
</table>

$^a$ See “Other scales.”
Fatigue Impact Scale (FIS) [58]

The FIS, also called the Fisk Fatigue Severity Score (FFSS), seeks to assess the impact of fatigue on different areas of functioning (cognitive, physical and psychosocial) rather than fatigue severity or phenomenology. It has good internal consistency and correlates \( r = .51 \) with the Sickness Impact Profile (a measure of general health status based on a patient’s description of how their functioning has been affected by their disease). The FIS was validated in a sample of patients with MS and hypertension, and significant differences were found in the scores of these two groups of patients on all subscales.

The FIS is an effective tool for assessing the impact of fatigue on patients’ lives. Validation in primary biliary cirrhosis patients has shown good reproducibility, suggesting that the scale could be of use in intervention trials [59]. The wording does assume that the patient is suffering from fatigue (because of my fatigue...), but this also allows a measure of attribution. The Daily Fatigue Impact Scale (DFIS) [60] has been developed from the FIS to assess daily changes in fatigue. Validated in patients suffering from flu-like illness, it has good internal consistency, construct validity and sensitivity to change.

Number of citations: 36.
Examples of use: MS [61,62], primary biliary cirrhosis [59,63], stroke [64] and brain injury [18].

Fatigue Rating Scale (FRS)
See FQ.

Fatigue Scale (FS)
See FQ.

Fatigue Questionnaire [65]

Also referred to as the FRS, the Chalder Fatigue Scale and the FS, this scale was developed for hospital and community studies of patients with CFS and has been used in this population in many studies since (first published in Ref. [66]). The FQ consists of 11 items measuring fatigue-related symptoms and loading onto two dimensions — physical and mental fatigue. This structure has been replicated in subsequent studies [67,68]. The scale was validated
against a fatigue item in the CIS, and a cutoff of 3/4 is recommended for identifying significant fatigue.

The scale has good clinical validity supported by a population study of fatigue in the general population [68]. FQ scores were continuously distributed with higher scores seen amongst those receiving disability allowances and those reporting disease and current health problems. The validity of the FQ in assessing fatigue in the general population suggests that it is a useful tool for assessing fatigue in a variety of medical disorders, although the presence of primary physical or cognitive dysfunction may confound interpretations of the responses. It has been used to assess fatigue in patients with conditions such as cancer and HIV and in general medical patients and Gulf War

<table>
<thead>
<tr>
<th>Scale name</th>
<th>Cancer Fatigue Scale</th>
<th>CIS</th>
<th>FACES⁴</th>
<th>FAI</th>
<th>FDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is assessed?</td>
<td>Phenomenology and severity</td>
<td>Phenomenology and severity</td>
<td>Phenomenology and severity</td>
<td>Phenomenology, severity, impact and possible triggers</td>
<td>Phenomenology, severity and frequency</td>
</tr>
<tr>
<td>Number of items</td>
<td>15</td>
<td>20</td>
<td>50</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>Scale type</td>
<td>5-point Likert</td>
<td>7-point Likert scale</td>
<td>4-point Likert scale</td>
<td>7-point Likert</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Number of subscales or factors</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Target population</td>
<td>Cancer</td>
<td>CFS</td>
<td>Conditions with associated energy-deficient states and research</td>
<td>General medical</td>
<td>MS</td>
</tr>
<tr>
<td>Standardisation sample(s) (n)</td>
<td>Cancer (307)</td>
<td>CFS (298)</td>
<td>Severe insomnia (372)</td>
<td>Lyme disease, CFS, post-Lyme chronic fatigue, SLE, MS and dysthymia and controls (235)</td>
<td>MS (155)</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>0.88</td>
<td>0.90</td>
<td>0.72–0.97</td>
<td>0.70–0.91</td>
<td>–</td>
</tr>
<tr>
<td>Test–retest reliability</td>
<td>0.69</td>
<td>–</td>
<td>Nk</td>
<td>0.29–0.69</td>
<td>–</td>
</tr>
<tr>
<td>Concurrent validity</td>
<td>VAS-F</td>
<td>Maslach Burnout Inventory-General Survey (MBI-GS) [124] exhaustion subscale</td>
<td>Sleepiness and consciousness scales with the Epworth Sleepiness Scale [125] and all scales with number of reported sleep problems</td>
<td>Subscale 1 with RAND Vitality Index [126], subscale 3 weakly with Enervation Scale [127]</td>
<td>FSS</td>
</tr>
<tr>
<td>Discriminative validity</td>
<td>Detects fatigue from nonfatigue in cancer population</td>
<td>Discriminates amongst CFS patients or MS patients, healthy controls and different occupational groups</td>
<td>–</td>
<td>Discriminates between patients and controls and some differences between patient groups</td>
<td>–</td>
</tr>
<tr>
<td>Cutoff score(s)</td>
<td>18/19</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>Yes</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

⁴ See “Other scales.”
veterans, although there is limited information as to its validity in these groups. It owes its popularity to its efficacy and the ease and speed with which it can be completed.

Number of citations: 174.

Examples of use: CFS [69–71], HIV [72], general population [68], cancer [73], Gulf War veterans [74] and MS [75,76].

**Fisk Fatigue Severity Score**
See FIS.

<table>
<thead>
<tr>
<th>FIS, FFSS</th>
<th>HRFS*</th>
<th>FQ (FRS, CFS, FS)</th>
<th>FSCL*</th>
<th>FSI</th>
<th>MAF, GFI</th>
<th>MFI-20, MFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact</td>
<td>Intensity, impact, aspects related to fatigue</td>
<td>Severity</td>
<td>Phenomenology</td>
<td>Severity, impact and duration</td>
<td>Severity, impact, distress and timing</td>
<td>Phenomenology severity and impact</td>
</tr>
<tr>
<td>40</td>
<td>56</td>
<td>11</td>
<td>30</td>
<td>13</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>5-point Likert</td>
<td>Likert</td>
<td>Yes/no response or 4-point Likert</td>
<td>Checklist</td>
<td>11-point Likert</td>
<td>100-mm visual analogue, later changed to 10-point Likert</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>MS</td>
<td>HIV</td>
<td>CFS</td>
<td>Nonclinical</td>
<td>Cancer</td>
<td>Rheumatoid arthritis</td>
<td>General medical</td>
</tr>
<tr>
<td>MS and hypertension (105)</td>
<td>Nonhospitalised HIV (54)</td>
<td>Primary care (374)</td>
<td>–</td>
<td>Women who had received or who were undergoing treatment for breast cancer and women without cancer (270)</td>
<td>Rheumatology clinic attenders</td>
<td>Cancer and CFS patients, healthy subjects (1423)</td>
</tr>
<tr>
<td>0.93</td>
<td>0.94</td>
<td>0.88–0.90</td>
<td>–</td>
<td>&gt;0.94</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td>–</td>
<td>0.43</td>
<td>–</td>
<td>0.35–0.75 (clinical), 0.10–0.74 (controls)</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Sickness Impact Profile (SIP) [128]</td>
<td>Revised Clinical Interview Schedule [129] (CIS-R) fatigue question</td>
<td>POMS-F and SF-36-vitality</td>
<td>POMS-F, and POMS-V</td>
<td>VAS-F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant difference between scores of MS and hypertensive patients on all scales</td>
<td>Discriminates between patients with and without fatigue assessed on CIS</td>
<td>Sensitive to fatigue in both breast cancer population and in a noncancer population</td>
<td>Detects significant differences in fatigue between patients and controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td>3/4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

**Fatigue Severity Inventory**
See FAI.

**Fatigue Symptom Inventory (FSI) [77]**
This multidimensional scale measures, in addition to severity, the duration of fatigue and its impact on quality of life in cancer patients. The initial standardisation sample consisted of women undergoing treatment for breast cancer, those who had completed breast cancer treatment and those who had never been diagnosed with cancer. It was shown to...
have moderately good psychometric properties, although the test–retest reliability was, in the authors’ own words, weak to moderate.

The scale has since been validated in a different cancer conditions [78] consisting of both male and female patients including the elderly, although there remains no evidence on its sensitivity to change over time or with treatment.

Overall, however, the scale’s validity in both genders, a wide age range and a variety of cancer diagnoses makes it a useful tool in assessing the impact and duration of fatigue. Although, to date, the scale has only been used in cancer studies, it reflects generic aspects of fatigue that may make it suitable for use in other populations.

Number of citations: 19.
Examples of use: cancer [78–80].

Fisk Fatigue Severity Score
See FIS.
Lee Fatigue Scale (LFS)

See Visual Analogue Scale for Fatigue (VAS-F).

Multidimensional Assessment of Fatigue (MAF) scale and the Global Fatigue Index (GFI) [81]

Although in principle a multidimensional scale, the MAF was designed to generate a single score, the GFI, from 15 items in five separate dimensions: Degree, Severity, Distress, Impact on activities of daily living and Timing [82]. A further item, not included in the GFI, measures change in fatigue over the past week. This instrument is unusual in that it allows patients to miss out irrelevant questions on the activities of daily living subscale.

Originally rated using visual analogue scales [81], later versions have employed 10-point numerical rating scales, still referred to as the MAF. In both forms, reasonable psychometric properties have been established although more information is needed as to the scale’s test–retest reliability and sensitivity to change over time. It has been found to be a valid and reliable measure for use in HIV [82], although its construct validity and appropriateness for use in cancer patients have been questioned [83].

Number of citations: 37.

Examples of use: rheumatoid arthritis [84,85], Cancer [83], MS [86]

Multidimensional Fatigue Inventory (MFI-20) [87]

This widely used measure consists of five subscales — General fatigue, Physical fatigue, Mental fatigue, Reduced motivation and Reduced activity. Internal consistency is good for all subscales as are test–retest reliability results (general fatigue $r = .83$, physical fatigue $r = .87$, reduced activity $r = .84$, reduced motivation $r = .80$ and mental fatigue $r = .74$) [88].

However, although it is one of the most comprehensive and promising fatigue measures currently available for use in cancer patients, it has been suggested that the scale needs further development before use in a clinical setting [83,87,89,90]. In the initial validation study, there were some surprising findings, e.g., the general fatigue scale did not discriminate between cancer patients and students, and students were found to have higher scores (i.e., more fatigue) than cancer patients on the mental fatigue scale. It also appeared that with the exception of the mental fatigue scale, all of the subscales behaved somewhat similarly, suggesting that the distinction between dimensions may not be as important as initially claimed. In a later test of the scale’s psychometric properties [83], a five-factor solution was obtained but with very different item loadings, which also suggests problems with the dimensional structure.

In a recent study, the MFI-20 was shown to discriminate between patients with and without Parkinson’s disease [55], although the contribution of non-fatigue-related parkinsonian motor and cognitive symptoms was not clear.

Number of citations: 99.

Examples of use: cancer [91–94], Sjögren’s syndrome [95], Parkinson’s disease [55], chronic obstructive pulmonary disease [96], rheumatoid arthritis [97].

Multidimensional Fatigue Symptom Inventory (MFSI) [89]

The MFSI assesses five dimensions of fatigue: Global experience, Somatic symptoms, Cognitive symptoms, Affective symptoms and Behavioural symptoms. The standardisation sample consisted of women who had received treatment for breast cancer and women who had no history of cancer.

The MFSI was found to have good psychometric properties. The scales factor structure shows a reasonable fit with the originally conceptualised dimensions, although different labelling is used (General fatigue, Emotional fatigue, Physical fatigue, Mental fatigue and Vigour). The MFSI has excellent internal consistency, good test–retest reliability, convergent validity and divergent validity for all dimensions. It also has diagnostic validity, with significant differences between scores of cancer patients and noncancer patients on subscales of General fatigue, Emotional fatigue, Physical fatigue and Vigour.

The authors suggest that, as the MFSI contains no reference to any medical diagnosis or disease, it may well be of use in assessing fatigue in other clinical and healthy populations and for making baseline assessments in patients about to undergo treatment that may cause fatigue. With appropriate validation, the MFSI is a potentially valuable tool in both research and clinical settings, although its length may limit its usefulness.

Number of citations: 16.

Examples of use: none to date.

Piper Fatigue Scale (PFS) [98]

The PFS, developed for use in research in cancer patients, has received various criticisms. It takes a long time to complete and patients have had difficulty understanding it. In addition, the wording assumes that the patient is already suffering from fatigue, requiring initial screening before use.

In terms of psychometric qualities, the original version has some shortcomings. Firstly, factor analytical techniques were not used to establish the validity of the dimensional structure. Secondly, the scale was validated on only 42 patients. Finally, although the internal consistency is high, concurrent validity measures are only moderate. In fact, when used together with the Fatigue Symptom Checklist (FSCL) (see “Other scales”), the only correlations found were with mood-related items on the PFS, while total PFS fatigue score did not correlate with any of the items on the FSCL.

Number of citations: 64.

Examples of use: cancer [99], HIV [100], chronic obstructive pulmonary disease [101] (found to be unsuitable) and well women [102].

Revised PFS [103]

In 1998, the Revised PFS was developed and validated in a sample of women recovering from breast cancer.
Factor analysis revealed four dimensions—Sensory, Affective meaning, Cognitive/mood and Behavioural/severity—and a number of redundant items were deleted. The response format was also changed to a Likert scale, making it easier to score. The internal consistency of the new scale is high, and a recent study has found good psychometric properties in a population of postpoliomyelitis patients, including high concurrent validity with the FQ ($r = .80$) and good test–retest reliability results ($r = .98$). Confusingly, this new version is still referred to as the PFS in most reports.

Number of citations: 14.

Examples of use: older adults [104] and postpoliomyelitis infection [105].

Schwartz Cancer Fatigue Scale (SCFS) [106]

The scale was developed for measuring cancer-related fatigue. Factor analysis revealed a four-factor solution that accounted for 70% of the variance. The factors were named Physical, Emotional, Cognitive and Temporal.

The 28-item scale is easy to administer and its psychometric properties appear to be good. However, there is no information as to its test–retest reliability, and further studies need to be carried out on larger samples to confirm its diagnostic and discriminatory ability. A more recent study in cancer patients has suggested a two-factor structure rather than the four-factor structure originally proposed [107].

Number of citations: 12.

Examples of use: cancer [107].

Visual analogue ratings of physical energy (PE) and mental energy (ME) [108]

These are simple, well-validated visual analogue scales (from $0 = \text{I have no energy at all}$ to $100 = \text{I am full of energy}$), which are quick to complete and allow patients to give different ratings for mental and physical dimensions.

Number of citations: 7.

Examples of use: healthy volunteers [109].

Visual Analogue Scale for Fatigue [110]

The VAS-F was designed to be a simple and quick measure of fatigue and energy levels for patients in the general medical population. As the name suggests, it comprises a number of visual analogue scales organised into energy and fatigue dimensions. The psychometric properties are good, although as concurrent validity was established using the Stanford Sleepiness Scale (SSS) [111], it has been suggested that the VAS-F scale is unable to distinguish between fatigue and sleepiness [18]. Similarly, it has been found sensitive to morning and evening changes in cancer patients [83]. It is sometimes called the LFS.

Number of citations: 36.

Examples of use: HIV [100], cancer [83,112,113], brain injury [18] and stroke [114].

Other scales

The measures considered so far have all been widely used or, if new, have provided sufficient evidence to evaluate aspects of their reliability, validity and utility. A number of other scales have also been reported. These are considered briefly for completeness and because they may be the subject of future use and evaluation. Kirsh et al. [115] investigated a single-item screening measure, “I get tired for no reason.” It has not been validated against a complete existing fatigue scale, and while it may prove useful as a brief screening measure, it has the same shortcomings as other single-item scales (see Ref. [39]). The Profile of Fatigue-Related Symptoms (PFRS) [116] is a multidimensional measure of symptoms associated with CFS rather than fatigue itself. The Cancer-Related Fatigue Distress Scale (CRFDS) [117] has good reliability but has not been validated in any other studies. The Swedish Occupational Fatigue Inventory (SOFI) [118], the Cancer Fatigue Scale [113] and the FSCL [119] have not been validated in English-speaking populations. The SOFA and the FSCL were developed for occupational groups, and while the FSCL has been validated in cancer patients, the SOFA has been shown not to be a valid instrument in a clinical population (also cancer patients). A new scale for HIV, the HIV-Related Fatigue Scale (HRFS) [120], has been developed by assembling items from a number of other existing measures, although it has not yet been adequately evaluated. The Fatigue, Anergia, Consciousness, Energized and Sleepiness Adjective (FACES) checklist [121] is a new 50-item multidimensional tool designed to characterise different qualities of fatigue/sleep states across conditions. It has been validated only in a sleep disorder population, raising the possibility that it is measuring constructs such as tiredness or sleepiness rather than fatigue. Finally, the Fatigue Descriptive Scale (FDS) [122] for MS correlates well with the FSS, but no other psychometric information is available.

Recommendations

Fatigue assessment depends on a clear understanding of the phenomenology and aetiology of fatigue within a condition. In developing fatigue scales, there is a “catch-22” situation: before a concept can be measured, it must be defined, and before a definition can be agreed, there must exist an instrument for assessing phenomenology. There is unfortunately no “gold standard” for fatigue, nor is there ever likely to be.

There are a number of issues to be considered in choosing a particular scale for research or clinical practice.

1. What aspects of fatigue are to be assessed and why?

As discussed above, the titles given to scales can be misleading. Careful examination of the scale items and evidence of convergent validity with other scales should be undertaken to ensure that the instrument is measuring the
core concept intended by the investigator. For example, some scales may be measuring tiredness or sleepiness rather than a more typical fatigue experience.

No two scales measure exactly the same thing. Some measure phenomenology and others measure fatigue severity or impact, while many assess a mixture of all these. Choice of scale is dependent on what aspect(s) of fatigue the clinician/researcher wishes to measure. It is also important to consider the purpose of the assessment. Where a scale is to be used to screen for, or diagnose, fatigue in individual patients or groups of patients, the instrument should have a proven ability to discriminate cases from noncases, with acceptable levels of sensitivity and specificity. For other studies seeking to describe fatigue severity or impact, the scale must be sensitive to the full range of presentations. Thus, while a brief instrument with a handful of items may be sufficient for a screening test or use as part of an epidemiological study, detailed investigations into fatigue or the measurement of change would require longer and more detailed questionnaires. Finally, where a scale is to be used as an outcome measure in a clinical trial, it should have proven sensitivity to change with disease progression or treatment.

2. Should you choose a unidimensional or multidimensional scale?

Unidimensional scales are designed to derive a single score that captures heterogeneous symptoms and behaviours. Such scales are often relatively brief, which makes them easy and more economical to administer and score, and therefore useful as outcome measures in large studies or as screening instruments. Where well constructed, unidimensional scales can show good levels of internal consistency and test–retest reliability. Multidimensional scales, on the other hand, are typically longer but provide a detailed qualitative and quantitative assessment of fatigue. This can make them useful for comparing profiles across conditions for descriptive research or in seeking to identify mechanisms underlying specific aspects of fatigue. However, the validity of individual subscales may vary, with some (particularly those with only a few items) having unacceptable reliability.

3. Is the scale suitable for use in your patient population?

Most scales have been designed for use in specific populations. Ideally, such scales will have been validated, e.g., in their ability to distinguish between cases and noncases and between different severities of fatigue within that population or to be sensitive to change following treatment or other intervention. Where there is no information on the validity of a particular scale in the target population/condition, the clinician or researcher may choose to use a scale designed for use in other populations; indeed, there are many instances of this in the literature. However, in the absence of independent validation, these studies should be interpreted with caution.

At its simplest level, the validity of a scale can be assessed in terms of its content. When using a scale in a population other than that for which it was developed, it is important to examine the individual items to assess any overlap between fatigue-related and non-fatigue-related symptoms. Fatigue in the normal population has symptoms relating to physical and cognitive function and interacts with depression and anxiety [123]. These same problems may occur as a consequence of the disease in the absence of a subjective experience of fatigue. A test score may therefore confound aspects of disease symptomatology with fatigue severity. Particular problems arise where the researcher or clinician wishes to assess fatigue in children or older adults where validity data are usually lacking.

Finally, patients with fatigue may have problems completing long questionnaires, particularly when the fatigue measure is part of a larger assessment pack. The scale chosen may therefore need to be a compromise between practicality in administration and level of detail obtained.

Conclusion

There is clearly much to be done in the development of new scales and in the further validation of those already in existence. Even basic data on reliability are missing on many scales; few provide evidence on sensitivity to change or suggest cutoff scores for identifying levels of clinical caseness. This latter shortcoming is particularly significant given the prevalence of fatigue within the general population. Although different scales are often used for cross validation, there have been almost no direct comparisons between the properties of different scales for specific purposes. Finally, few scales have attended to the questions of possible age, sex, ethnic, educational, cultural and socioeconomic factors. Given the wide range of mechanisms probably underlying fatigue, differing manifestations and confounding effects of disease symptoms and/or treatment, it seems unlikely that any one fatigue scale will ever be appropriate for measuring fatigue in all disease groups. It is hoped that the present review will provide guidance on choosing between available scales and highlight the need for the development and validation of effective generic and disease-specific measures.

Acknowledgments

This work was supported by a grant from the Parkinson’s Disease Society (UK).

References


