review

A systematic review of the scales used for the measurement of cancer-related fatigue (CRF)

O. Minton* & P. Stone
Division of Mental Health, St George's University of London, London, UK

Received 9 April 2008; revised 16 May 2008; revised 27 June 2008; accepted 30 June 2008

Background: Fatigue in cancer is very common and can be experienced at all stages of disease and in survivors. There is no accepted definition of cancer-related fatigue (CRF) and no agreement on how it should be measured. A number of scales have been developed to quantify the phenomenon of CRF. These vary in the quality of psychometric properties, ease of administration, dimensions of CRF covered and extent of use in studies of cancer patients. This review seeks to identify the available tools for measuring CRF and to make recommendations for ongoing research into CRF.

Methods: A systematic review methodology was used to identify scales that have been validated to measure CRF. The inclusion criteria required the scale to have been validated for use in cancer patients and/or widely used in this population. Scales also had to meet a minimum quality score for inclusion.

Results: The reviewers identified 14 scales that met the inclusion criteria. The most commonly used scales and best validated were the Functional Assessment of Cancer Therapy Fatigue (FACT F), the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ C30) (fatigue subscale) and the Fatigue Questionnaire (FQ).

Conclusions: Unidimensional scales are the easiest to administer and have been most widely used. The authors recommend the use of the EORTC QLQ C30 fatigue subscale or the FACT F. The FQ gives a multidimensional assessment and has also been widely used. A substantial minority of the scales identified have not been used extensively or sufficiently validated in cancer patients and cannot be recommended for routine use without further validation.

Key words: cancer-related fatigue, measurement scales, systematic review

introduction

Cancer-related fatigue (CRF) is a subjective symptom experienced by patients at all stages of disease. It can occur during treatment [1], in advanced disease [2] and in disease-free survivors [3]. However, the prevalence of fatigue can vary widely depending on which measurement tool is used [4]. This is in part a reflection of the lack of an agreed definition of CRF. Diagnostic criteria for a syndrome of CRF have been proposed for inclusion in the International Classification for Diseases (ICD 10) (Cella ref). These criteria can be applied using a short semistructured interview [5]. The widespread adoption of a syndrome approach using agreed diagnostic criteria would help with the interpretation of prevalence figures in different populations and across different studies. However, the syndrome approach has not yet been widely used, probably because a strict application of the criteria requires the use of a semistructured psychiatric interview to ensure that symptoms of CRF are not related to underlying psychiatric co-morbidity.

The European Association of Palliative Care has developed a working definition [6] of CRF on the basis of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ C30). This defines fatigue as a subjective feeling of tiredness, weakness or lack of energy. This is a pragmatic approach which provides a working understanding of CRF for clinicians. However, other organisations and authorities have defined CRF in their own ways, e.g. the National Comprehensive Cancer Network [7], and there is no universally accepted definition.

The lack of consensus in this area has led to the development of a number of scales to measure CRF. These scales have usually been validated originally in cancer patients. Other investigators have used scales that were originally validated in non-cancer...
populations and have then independently validated their use in patients with cancer. These scales can differ widely in the number of items that they contain, the dimensions of CRF that they cover (e.g. physical, affective and cognitive) and their psychometric properties.

In the light of this multiplicity of different assessment methods, we undertook a review to identify which scales have been best validated and to make recommendations about which instruments should be used in research and/or in routine clinical practice.

**Methods**

A systematic methodology was used. The following terms were used to search Medline, CINAHL and PsychINFO (1950—week 1 February 2008):

1. exp "Outcome Assessment (Health Care)/"
2. exp Psychometrics/
3. exp "Outcome and Process Assessment (Health Care)/"
4. (fatigue adj (scale or inventory or instrument or measurement or assessment)).mp.
5. (fatigue adj2 (scale or inventory or instrument or measurement or assessment)).mp.
6. (fatigue adj3 (scale or inventory or instrument or measurement or assessment)).mp.
7. (fatigue adj4 (scale or inventory or instrument or measurement or assessment)).mp.
8. exp Questionnaires/
9. 9 or/1-8
10. exp Fatigue/
11. fatigue ((lab,kw)
12. ((lack$ or loss or lost) adj2 (energy or vigour or vigor)).mp.
13. (tired$ or weary or weariness or exhaustion or exhausted or lacklustre or astheni$).mp.
14. (apathy or apathetic or lassitude or letharg$ or (feeling adj3 (drained or sleepy or sluggish or weak4))).mp.
15. or/10-14
16. 9 and 15

The titles and abstracts of the papers identified using this search strategy were screened by one of the authors (O.M.) and where necessary the full text articles were retrieved. The reference lists of included articles were also examined. O.M. was responsible for the search strategy and retrieving studies. An ad hoc quality score was created before searching, to ensure that included scales reached a minimum level of psychometric properties on the basis of the ideal characteristics of measurement scales described in detail by Norman & Streiner [6, 7]. These characteristics are detailed in Table 1. The final list of included studies was agreed by both authors.

Scales were included only if they met the following a priori criteria:

1. Self-assessment scales originally validated in cancer patients and/or subsequently widely used in cancer populations (n > 50 patients).
2. Scales must have been used in a second test population for independent validation of their use in cancer patients.
3. 90% of participants must be ≥18 years.
4. Scales obtained a minimum quality score. The paper describing the scale must have contained information on at least three of the following:

**Table 1. Explanation of psychometric properties of an ideal scale (on the basis of Norman & Streiner [6, 7])**

<table>
<thead>
<tr>
<th>Psychometric evaluation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Do items within the scale correlate with each other and with the total score? This is usually measured by Cronbach’s alpha. A low value indicates a poor consistency of items; a very high score indicates item redundancy.</td>
</tr>
<tr>
<td>Test–retest reliability</td>
<td>Does the scale provide a similar score when administered to the same population in the same setting at different times?</td>
</tr>
<tr>
<td>Discriminant validity</td>
<td>Does the scale distinguish between groups expected to have differing levels of fatigue, e.g. cancer patients and a control population.</td>
</tr>
<tr>
<td>Responsiveness to change</td>
<td>Does the scale measure change during the course of an intervention that alters the level of fatigue, e.g. during chemotherapy.</td>
</tr>
<tr>
<td>Convergent validity</td>
<td>Does the scale provide a similar measure to previously validated fatigue tools? This is usually measured by the degree of correlation of the scores. The value usually given is Pearson’s r correlation.</td>
</tr>
</tbody>
</table>

Scales were included only if they met the following a priori criteria:

5. Scales must be in English or translated and validated for English language use.

There were also explicit exclusion criteria:

1. Objective rating scales testing power/strength, etc. as opposed to subjective fatigue.
2. Single-item scales, including visual analogue scale (VAS)
3. Fatigue subscales as part of a broader quality-of-life measure. Unless specific data were available relating to the psychometric properties of the subscale.

In addition, the original reference was cross-referenced for citing articles via Medline and Web of Knowledge. This was carried out to quantify the number of times the scales had been used and the type of populations studied. This procedure allowed us to approximately quantify the number of participants who had been tested with each scale. For ease of reference, populations of cancer patients were divided into three groups:

1. Active—on treatment with curative intent.
2. Palliative—on treatment with palliative intent or supportive care only.
3. Disease free—medium to long-term survivors.

This methodology allowed us to make firmer recommendations for ongoing research into CRF by being able to compare different levels of scale usage as well as the scales’ psychometric properties.

**Results**

The search strategy identified 7889 abstracts in total. These were screened and 116 studies were identified where the main focus of the paper was on fatigue measurement. This
shortlist generated data on 22 different scales that have been used in the assessment of CRF (see Figure 1). Eight of these scales did not meet our minimum quality requirements. The specific reasons for excluding these instruments are itemised below.

The Fatigue Symptom Control Checklist [9]: Although used in some studies of CRF (e.g. by Morrow et al. [10]), this scale has neither been validated in English nor been specifically validated for use in cancer patients.

The Swedish Occupational Fatigue Inventory [11]: This scale has been used in one study of patients undergoing radiotherapy [12] but was found to have limitations and has not been further evaluated.

The original Piper Fatigue Scale (PFS) [13]: This scale was not included as it was found to have a number of flaws in its use. It also required the use of an initial screening tool to identify patients with fatigue. The revised version of the scale was included (see below).

The Cancer Fatigue Scale [14] and the Fatigue Assessment Questionnaire [15] were both excluded as validated English language versions have not been published.

The Fatigue Management Barriers Questionnaire [16], the Clinical Survey for CRF (QFAS) [17] and the Cancer-Related Fatigue Distress scale [18] were all excluded because although they have undergone pilot evaluation in cancer patients no further published reports regarding their validity could be identified.

This left 14 scales which met our criteria and have been included in the review. The scales have been divided into unidimensional and multidimensional for ease of presentation. The unidimensional scales invariably cover the physical aspects of fatigue only. The multidimensional scales cover anywhere between two and five different aspects of fatigue. This division is possibly artificial as CRF has been shown in a recent study to be a unidimensional construct, even after measuring all relevant dimensions of fatigue [19]. However, this method of categorising measurement scales is still widely used and has been adopted in this review in order to structure the presentation and discussion of the different instruments.

14 scales included
8 excluded on quality criteria
5 unidimensional scales
9 multidimensional scales

Figure 1. Flow chart of study identification and selection.

unidimensional scales

We found five scales that reached our minimum quality standard. Full details of their psychometric properties and areas of use are detailed in Table 2.

The Brief Fatigue Inventory (BFI) [23] is a nine-item VAS that was validated for use in a mixed cancer population. It has reasonable psychometric properties but has had limited ongoing use. The scale has cut-off scores to differentiate between mild, medium and severe fatigue but these have not been validated and are likely to be of use for screening purposes only.

The EORTC QLQ C30 [24] is a 30-item quality-of-life questionnaire. The full tool has been used extensively as a quality-of-life instrument. The three-item fatigue subscale has been independently validated as a separate fatigue measure. While the psychometric properties are weaker than more extensive scales, their brevity and ease of administration may outweigh this disadvantage. There have also been two large-scale studies (>2000 patients in each) independently assessing its use [20, 21], so there are extensive data in a variety of settings. However, it has been noted to have a ceiling effect in advanced cancer patients [20] and is not recommended as a single measure in this group.

Another advantage of this measure is that it is possible to extract fatigue data from studies that have used the full quality-of-life scale. This includes the large number of studies that have used this scale in cancer chemotherapy trials (note: these studies have not been included in the current analysis).

The Fatigue Severity Scale [25] is a nine-item scale that was originally validated in a chronic illness (non-cancer) population and while it has been extensively used in neurological disease and chronic fatigue, it has had very limited use in cancer patients. It has only been used and validated by one author in a few related studies [1, 26]. However, no other studies were identified and it is not recommended for use in the measurement of CRF.

The Functional Assessment of Cancer Therapy Fatigue (FACT F) subscale [27] is a 13-item stand-alone scale but is one of a range of the FACIT series of quality-of-life and tumour-specific symptom questionnaires [22]. It has been used in a number of intervention studies to treat CRF, and the original authors have been able to derive change in scale scores that correspond to minimum clinically significant differences [28]. This application makes it an especially useful measure for intervention studies.
The Profile of Mood States [29] was originally used as a measure of workforce health. It contains a number of scales including a fatigue subscale of seven items which has been independently examined in both a non-cancer [30] and a cancer population [31]. It has also been used to provide convergent validity in the validation of a number of other scales used in CRF. It has a defined minimum clinically significant difference [32]. Its extensive use in studies has meant there are ample data available to recommend its continued use in this setting.

### Multidimensional scales

There were nine scales included in this category. Full details of these scales can be found in Table 3.

### Table 2. Unidimensional scales

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Scale type</td>
<td>Numerical</td>
<td>Likert</td>
<td>Numerical</td>
<td>Numerical</td>
<td>Likert</td>
</tr>
<tr>
<td>Population studied in original validation study</td>
<td>Mixed cancer</td>
<td>Three populations: lung cancer, bone marrow transplant, metastatic cancer</td>
<td>Chronic illness population</td>
<td>Mixed cancer patients on treatment</td>
<td>Mixed work population and psychiatric patients</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>0.96</td>
<td>0.80–0.85</td>
<td>0.96</td>
<td>0.95</td>
<td>0.90–0.94</td>
</tr>
<tr>
<td>Dimensions fatigue covered</td>
<td>Physical functioning</td>
<td>–</td>
<td>Physical functioning</td>
<td>Physical functioning</td>
<td>Physical functioning</td>
</tr>
<tr>
<td>Test &amp; retest reliability</td>
<td>–</td>
<td>–</td>
<td>Subset 87 patients, 3/52 later measurement, error 4.7 units</td>
<td>r = 0.90 (3–7 days later)</td>
<td>r = 0.66 (12–16 weeks later)</td>
</tr>
<tr>
<td>Known group comparison (discriminant validity)</td>
<td>Distinguished CRF from community adults</td>
<td>Distinguished between local and regional disease</td>
<td>Distinguished CRF versus healthy controls</td>
<td>Distinguished between three known haemoglobin levels (&gt;13, 11–13 and &lt;11)</td>
<td>Distinguished between chronic fatigue syndrome and controls (N = 6275)</td>
</tr>
<tr>
<td>Responsiveness to change</td>
<td>Not done</td>
<td>On the basis of performance status deterioration</td>
<td>Not done</td>
<td>Minimum clinical important difference (MCID) derived (3 points)</td>
<td>MCID derived (5.6 points)</td>
</tr>
<tr>
<td>Convergent validity (against other scales)</td>
<td>POMS F, r = 0.84</td>
<td>Fatigue Questionnaire, r = 0.49–0.75</td>
<td>EORTC, r = 0.83</td>
<td>Piper Fatigue Scale, r = 0.75; POMS F, r = 0.74</td>
<td>Not done</td>
</tr>
<tr>
<td>Number of participants tested in original validation study</td>
<td>305</td>
<td>366</td>
<td>227</td>
<td>50</td>
<td>695</td>
</tr>
<tr>
<td>Populations studied (active/palliative/survivors)</td>
<td>A/P/S</td>
<td>A/P/S</td>
<td>A/P</td>
<td>A/P/S</td>
<td>A/P/S</td>
</tr>
<tr>
<td>Approximate number of further cancer studies and participants*</td>
<td>4 (~2300) &gt;10 (&gt;5000)</td>
<td>1 (100) &gt;50 (&gt;5000)</td>
<td>&gt;20 (&gt;2000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where internal consistency—Cronbach’s alpha unless otherwise stated; convergent validity—Pearson’s correlation value unless otherwise stated. A, active treatment population; P, palliative care population; S, disease-free survivors; CRF, cancer-related fatigue. Scale abbreviations: BFI, Brief Fatigue Inventory; EORTC QLQ C30 FS, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Fatigue subscale; FSS, Fatigue Severity Scale; FACT F, Functional Assessment of Cancer Therapy Fatigue subscale; POMS F, Profile of Mood States Fatigue subscale.

*This involved cross-referencing studies in Medline and Web of Science and identifying studies in cancer where the scale has been used subsequently.

The Profile of Mood States [29] was originally used as a measure of workforce health. It contains a number of scales including a fatigue subscale of seven items which has been independently examined in both a non-cancer [30] and a cancer population [31]. It has also been used to provide convergent validity in the validation of a number of other scales used in CRF. It has a defined minimum clinically significant difference [32]. Its extensive use in studies has meant there are ample data available to recommend its continued use in this setting.

**Multidimensional scales**

There were nine scales included in this category. Full details of these scales can be found in Table 3.

The Chalder Fatigue Scale—also known as the Fatigue Questionnaire (FQ) [33]—is an 11-item scale that was originally validated in a general practice setting. However, its main use has been in the investigation of chronic fatigue syndrome. It is brief and easy to administer while still covering two aspects of fatigue (mental and physical). It has been used in population studies and so has normative data available for comparison with cancer patients [34, 41].

The Fatigue Symptom Inventory (FSI) [35] is a 13-item scale that was originally validated in a breast cancer population. A subsequent validation study involved 342 mixed cancer patients [42]. It has reasonable psychometric properties but there is a question over its test–retest reliability. It has been used in a number of studies but with overall small numbers of patients.
Table 3. Multidimensional scales

<table>
<thead>
<tr>
<th>Scale name</th>
<th>Number of items</th>
<th>Scale type</th>
<th>Original validation sample</th>
<th>Internal consistency</th>
<th>Dimensions fatigue covered</th>
<th>Test &amp; retest reliability</th>
<th>Known group comparison (discriminant validity)</th>
<th>Responsiveness to change</th>
<th>Convergent validity (against other scales)</th>
<th>Number of participants tested in original sample</th>
<th>Populations studied (active/palliative/survivors)</th>
<th>Approximate number of further cancer studies and participants</th>
<th>Scale name</th>
<th>Number of items</th>
<th>Scale type</th>
<th>Original validation sample</th>
<th>Internal consistency</th>
<th>Dimensions fatigue covered</th>
<th>Test &amp; retest reliability</th>
<th>Known group comparison (discriminant validity)</th>
<th>Responsiveness to change</th>
<th>Convergent validity (against other scales)</th>
<th>Number of participants tested in original sample</th>
<th>Populations studied (active/palliative/survivors)</th>
<th>Approximate number of further cancer studies and participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>PQ [30]</td>
<td>11</td>
<td>Likert</td>
<td>General practice</td>
<td>0.88–0.90</td>
<td>Physical and mental</td>
<td>Not done</td>
<td>Distinguished between Hodgkin’s disease and general population</td>
<td>Not done</td>
<td>Fatigue item on clinical interview schedule</td>
<td>POMS F, r = 0.75</td>
<td>A/P/S</td>
<td>&gt;10 (~2000)</td>
<td>MFI-20 [36]</td>
<td>20</td>
<td>Likert</td>
<td>Mixed cancer population</td>
<td>0.84</td>
<td>Cognitive, physical and emotional</td>
<td>Not done</td>
<td>Distinguished between cancer patients, healthy controls and doctors on shift work</td>
<td>Monitored change over course of radiotherapy</td>
<td>PFS (revised) r = 0.97</td>
<td>28</td>
<td>Schwartz CFS [39]</td>
<td>9</td>
</tr>
<tr>
<td>FSI [33]</td>
<td>13</td>
<td>Numerical</td>
<td>Breast cancer on treatment</td>
<td>0.94</td>
<td>Physical and mental</td>
<td>Range of values: patients, $r = 0.35–0.75$; controls, $r = 0.10–0.74$</td>
<td>Distinguished between healthy controls and breast cancer patients</td>
<td>Limited ability to detect changes over treatment period</td>
<td>Not done</td>
<td>fatigue item on clinical interview schedule</td>
<td>POMS F, r = 0.78</td>
<td>A/S</td>
<td>7 (~700)</td>
<td>MFSI-30 [37]</td>
<td>30</td>
<td>Likert</td>
<td>Breast cancer population</td>
<td>0.87–0.96</td>
<td>Cognitive, physical and mental</td>
<td>Not done</td>
<td>Distinguished between cancer patients and healthy controls</td>
<td>Not done</td>
<td>PFS (revised) r = 0.97</td>
<td>22</td>
<td>Schwartz CFS [39]</td>
</tr>
<tr>
<td>VASF [34]</td>
<td>18</td>
<td>Visual analogue scale</td>
<td>Patients with sleep disorders</td>
<td>0.91–0.96</td>
<td>Physical and mental</td>
<td>Not done</td>
<td>Distinguished between patients with sleep disorders and healthy controls</td>
<td>Not done</td>
<td>Fatigue item on clinical interview schedule</td>
<td>POMS F, r = 0.78</td>
<td>A only</td>
<td>122</td>
<td>PFS (revised) r = 0.97</td>
<td>22</td>
<td>Likert</td>
<td>Breast cancer population</td>
<td>0.97</td>
<td>Temporal, intensity, cognitive, affective and sensory</td>
<td>Not done</td>
<td>Distinguished between lung and breast cancer patients</td>
<td>Not done</td>
<td>PFS (revised) r = 0.97</td>
<td>22</td>
<td>Schwartz CFS [39]</td>
<td>28</td>
</tr>
<tr>
<td>MAF [35]</td>
<td>16</td>
<td>Numerical</td>
<td>Arthritis patients</td>
<td>0.93</td>
<td>Distress, interference, severity, cognitive</td>
<td>Not done</td>
<td>Distinguished between patients with arthritis and healthy controls</td>
<td>Not done</td>
<td>Fatigue item on clinical interview schedule</td>
<td>POMS F, r = 0.78</td>
<td>A only</td>
<td>50</td>
<td>Wu CFS [40]</td>
<td>9</td>
<td>Likert</td>
<td>Breast cancer population</td>
<td>0.91</td>
<td>Physical, emotional and cognitive</td>
<td>Not done</td>
<td>None</td>
<td>Not done</td>
<td>Wu CFS [40]</td>
<td>9</td>
<td>Wu CFS [40]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This involved cross-referencing studies in Medline and Web of Science and identifying studies in cancer where the scale has been used subsequently.

Its areas of measurement have been limited to patients undergoing active treatment and survivors.

The Lee fatigue scale (or Visual Analogue Scale for Fatigue—VASF) [36] is an 18-item tool that was originally validated in a group of patients with sleep disorders. It has had very limited use in cancer patients. Its psychometric properties were assessed in a sample of 212 mixed cancer patients undergoing treatment [31]. However, because of a potential overlap with measures of sleep disturbance it is not recommended for use in CRF measurement.

The Multidimensional Assessment of Fatigue (MAF) [43] is a 16-item scale that was originally validated in rheumatoid arthritis patients. Its psychometric properties were assessed in the same sample of cancer patients previously discussed with respect to the VASF [31]. It was also used (as one of a number of fatigue measures) by Morrow and colleagues in two subsequent intervention studies examining the role of antidepressants in treating fatigue [10, 37]. Compared with other fatigue instruments included in this review, it has been relatively poorly validated and its use in further studies cannot be recommended unless further validation work is undertaken.

The Multidimensional Fatigue Inventory (MFI-20) [44] is a 20-item scale that was designed for use in cancer patients. Its original validation had a number of group comparisons including both healthy controls and groups of subjects who at various time points were expected to be fatigued. This included army trainees and doctors undertaking shift work. This means there are normative data available for reference. It was further validated for use in cancer in a subsequent study of 275 patients undergoing radiotherapy [38]. It has since been used in a number of studies but with small numbers of patients in each study.

The Multidimensional Fatigue Symptom Inventory short form (MFSI-30) [39] is a 30-item scale that was originally investigated in a group of breast cancer patients undergoing treatment. It was further validated in a mixed cancer population of 304 patients [40]. It has favourable psychometric properties; however, its use in clinical studies has been somewhat limited. There have only been two additional studies reporting its use in breast cancer survivors.

Its use beyond this group will require further validation.

The revised PFS [45] is a shorter version of the original scale [13]. Piper reported some limitations to her original fatigue scale [13] and she went on develop a revised version. This revised scale consists of 22 items and has been validated in a group of breast cancer survivors [45]. There are limited data on the psychometric properties of the revised scale for use in cancer patients although data are available in other populations. While the scale has since been used in a small number of studies of cancer patients, the majority of these studies have been undertaken in breast cancer patients undergoing treatment and in breast cancer survivors. There are little or no data in other cancer populations.

The Schwartz Cancer Fatigue Scale [46] is a 28-item scale that was validated in a mixed cancer population undergoing treatment. Its psychometric properties were further examined (along with the VASF and the MAF) in the study of 212 mixed cancer patients discussed previously [31]. Schwartz has also determined the minimum clinically significant difference in a further study [32] on 103 mixed cancer patients undergoing treatment. We did not find any other studies that have used this scale. Its usefulness despite extensive psychometric data must therefore be questioned.

The Wu Cancer Fatigue Scale [47]: The original scale had 16 items but a secondary validation study found redundant items and the scale was revised to include only nine items [48]. The instrument has undergone limited psychometric evaluation and has not been used in any subsequent studies. The relative lack of data means that it cannot be recommended for further use.

**Discussion**

This review has demonstrated the range and number of scales available to specifically measure CRF. A number of other tools...
have been validated for the measurement of fatigue in nonmalignant conditions and in the general population but evaluation of these scales was beyond the scope of this review.

While this review has used systematic methodology, there are some potential limitations. The titles and abstracts were only screened by one author—this was because of limited time and resources. However, both authors agreed on the final included studies. This includes three scales which were identified but excluded from full analysis because only one published paper was available. These authors were not contacted for further unpublished data as a meta-analysis was not possible with this type of review. As a result, obtaining these additional data would not have materially altered the conclusions of the review (which concerned peer-reviewed published data only).

Our estimate of the number of studies that the various scales have been used in is on the basis of a comprehensive cross-referencing in both Medline and the ISI citation index with full text article retrieval if necessary. The purpose of including these approximate figures is to give the reader an ‘order of magnitude’ assessment about how often these scales have been used in previous studies. However, this method is not foolproof (e.g. authors may have failed to reference the scale correctly) and so this is stated to be approximate numbers only.

The unidimensional scales (which measure the physical impact of fatigue) are the most widely used. They also have some of the most robust psychometric data to support their use. Their limitation is the scope of measurement—subjective fatigue is more than the sensation of physical impairment. Although measuring physical fatigue symptoms may include the social and functional impact of physical fatigue. This extends beyond one aspect of fatigue that could imply that these scales cover more than one dimension of CRF.

However, any theoretical limitation is compensated for by their ease of use and brevity (the scales contain between three and 13 items). The most widely used of these scales are the FACT F and the EORTC QLQ C30 fatigue subscale which have data from over 10 000 patients between them and have been widely used in intervention studies to treat CRF [49]. The FACT F has the advantage of having a validated clinically significant score change—it also covers the social impact of CRF but takes longer to administer than the three-item EORTC QLQ C30 fatigue subscale. The former should be used in a research setting and the latter could be used to monitor clinical effects. The Profile of Mood States Fatigue has also been extensively used but was not originally validated in cancer patients and has no clear advantage over the other two scales. It has been used to help validate six out of the 14 scales included. The widespread use in a healthy population could provide a useful baseline measure of fatigue in the general population [50] for comparison with a cancer patient group.

The multidimensional scales are much more limited in their usage. While they offer the theoretical advantage of covering more aspects of fatigue, such as the cognitive or affective symptoms, this is often at the expense of an increase in the number of items. The more complex administration and completion time may be why their usage has been more limited. Moreover, the advantages of measuring additional ‘dimensions’ of fatigue are not clear [19]. As yet, there is no clinical value in distinguishing between patients with predominantly ‘physical’, ‘cognitive’ or ‘motivational’ fatigue.

Until the clinical importance of these proposed dimensions has been clarified there is little incentive to use these scales outside of a research study. Indeed, some authors suggest that this construct is redundant as CRF can be regarded as a unidimensional phenomenon [19]. It is perhaps relevant that a number of multidimensional scales have been developed that have not been subsequently extensively used. Some of these scales have up to 30 items and so could potentially more than double the administration time of all the unidimensional scales. The majority of the scales (seven out of nine) have data on <1000 patients for any individual scale. In addition, three of these seven scales have only been used in breast cancer patients and so their use is even more limited. The two exceptions are the FQ and the MFI-20. These two scales have each been used to measure CRF in >2000 patients. However, this evaluates still considerably fewer patients than the FACT F or EORTC QLQ C30. The FQ, by virtue of consisting of only 11 items, offers a two-dimensional (physical and mental) assessment of fatigue without an increase in the time required for administration. It was not, however, originally designed for use in cancer patients, whereas the MFI-20 was specifically created to measure fatigue in this population. Ultimately, the trade-off between undertaking a comprehensive multidimensional assessment of fatigue and the consequent problems of missing data and questionnaire ‘burden’ will be a decision for individual researchers.

conclusions

For most purposes, a unidimensional fatigue instrument is the appropriate measure. Of the many unidimensional scales available the best validated and most widely used are the EORTC QLQ C30 fatigue subscale and the FACT F. The EORTC subscale has the virtue of brevity and is usually used as part of the EORTC QLQ C30 quality-of-life assessment instrument. However, its extreme brevity and limited number of response categories may make it less sensitive to changes in fatigue and less able to detect differences in fatigue between groups. For this reason, the authors prefer the FACT F. This instrument is relatively brief, has robust psychometric properties and can also be easily combined with a validated quality-of-life assessment instrument (such as the FACT G). In those circumstances where a multidimensional fatigue instrument is required, the authors recommend the use of the FQ. Although not originally developed for use in cancer patients, this scale has robust psychometric properties and has been extensively used in other populations. It is brief, easy to use and its dimensions have face validity.

funding

OM was supported by a Wellcome Trust Value in People (VIP) award.
references


