Clinical Rehabilitation

<http://cre.sagepub.com/>

How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far

Sarah F Tyson and Philip Brown

Clin Rehabil published online 11 February 2014

DOI: 10.1177/0269215514521043

The online version of this article can be found at:

<http://cre.sagepub.com/content/early/2014/02/10/0269215514521043>

Published by:

<http://www.sagepublications.com>

Additional services and information for *Clinical Rehabilitation* can be found at:

Email Alerts: <http://cre.sagepub.com/cgi/alerts>

Subscriptions: <http://cre.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Feb 11, 2014

[What is This?](#)

How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far

Clinical Rehabilitation

1–13

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0269215514521043

cre.sagepub.com

Sarah F Tyson and Philip Brown

Abstract

Objective: To systemically review the psychometric properties and clinical utility (feasibility) of tools to measure fatigue in neurological conditions.

Data sources: AMED, CINAHL, MEDLINE, Science Direct and Web of Knowledge were searched for studies of the psychometric properties of fatigue measures in adults with central nervous system disorders. Search terms were: Fatigue AND assess* OR measure* OR tool OR outcome OR index OR test OR scale AND stroke OR cerebrovascular* OR hemi* OR parkinson* OR sclerosis OR head injury OR brain injury OR motor neurone disease OR neuro*

Review methods: Articles that addressed validity; reliability and ability to detect change were selected. Two authors independently selected articles and extracted data. Strength of the psychometric properties and clinical utility were assessed against predetermined thresholds.

Results: A total of 25 articles were selected and 17 measurement tools identified including six versions of the Fatigue Severity Scale and five versions of the Fatigue Impact Scale. All were clinically feasible but none met all the criteria. Most were valid. Reliability and ability to detect change were infrequently evaluated, but were adequate when reported. The most thoroughly developed and psychometrically robust measures were the Neurological Fatigue Indices.

Conclusion: Although many measures of fatigue in neurological conditions have published, none met all of the criteria for psychometric robustness and clinical utility but the Neurological Fatigue Indices are the best and can be used with caution

Keywords

Clinimetrics, fatigue, measurement, neurological disorders

Received: 17 May 2013; accepted: 29 December 2013

Stroke Research Centre, University of Manchester,
Manchester, UK

Corresponding author:

Sarah Tyson, Stroke Research Centre, School of Nursing,
Midwifery and Social Work, Jean McFarlane Building,
University of Manchester, Oxford Road, Manchester
M13 9PL, UK.

Email: sarah.tyson@manchester.ac.uk

Introduction

Although fatigue is an almost universal feature of neurological conditions, it is poorly understood and there are few treatment options.¹ Not surprisingly there are many challenges to managing fatigue, not least the lack of a widely accepted measurement tool that limits attempts to quantify the frequency, nature and severity of the problem. Good quality, standardized measurement tools that produce robust, accurate information and are practicable to use, are key to understanding the nature and severity of patients' problems and the evaluating effectiveness of interventions.^{2,3} Although the value of using such tools is recognized, clinicians report difficulty deciding which to choose⁴ and inconsistent use of outcome measures has hampered development of the evidence base for many years. Thus, as part of a wider programme of work to evaluate measurement tools for neurological rehabilitation, we systematically reviewed the psychometric properties and clinical utility (the feasibility of use) of measures of fatigue in neurological conditions to identify and recommend the most effective for use in research and clinical practice.

Several previous reviews have considered tools measuring neurological fatigue and have recommended a confusingly broad range of tools.⁵⁻⁸ A shortcoming is that only one⁸ systematically reviewed the methodological quality of the selected tools, making it difficult to judge the strength of the evidence to support their recommendations. This used the COSMIN checklist to evaluate methodological quality,⁸ which relies heavily on classic test theory.⁹ The review was also limited to measures for people with stroke, Parkinson's disease and multiple sclerosis, and the searches completed up to 2010. Thus, we reviewed tools measuring fatigue in all central nervous system disorders, systematically evaluated the psychometric properties and clinical utility of the tools and used contemporary searches.

Methods

Studies published in English were included if they assessed the psychometric properties of a tool to

measure fatigue in adults with a central nervous system disorder. The psychometric properties considered were:

- validity (concurrent or criterion related; construct or content);
- reliability (inter-rater or test re-test);
- ability to detect change.

The methods to assess these properties that we selected were:

- for concurrent or criterion-related validity: (parametric or non-parametric) correlation coefficients;
- for construct or content validity:
 - internal consistency (Cronbach's alpha) or factor/principal component analysis;
 - scaling properties – Rasch or Mokken analysis or co-efficient of scalability (for hierarchy); inter-item correlations or coefficients of reproducibility (for redundancy of items);
 - floor and ceiling effects;
- for reliability: intra-class correlations (for parametric data) and kappa statistics (for non-parametric data) or percentage agreement;
- for ability to detect change: measurement error, standardized response mean, standardized error of measurement, limits of agreement, minimal detectable change or ability to detect change during treatment.

Studies were excluded if they included measurement tools that were a composite of different constructs where the data for the fatigue-related items or sub-scales could not be extracted; measured psychometric properties or used methods other than those listed above; included a sample where less than 50% of the participants had a central nervous system disorder or the data for the relevant participants could not be extracted; involved patients who did not have a central nervous system disorder (such as peripheral neuropathy or chronic fatigue syndrome); related to a measurement tool in which only one psychometric property had been measured (as this would supply insufficient data to consider it for use); or was an instrumented

measure where the equipment could not be obtained (or the instructions to make it were unavailable).

The following electronic databases were searched; AMED, CINAHL, MEDLINE, Science Direct and Web of Knowledge from their inception to the beginning of April 2013 using the following keywords:

- Fatigue
AND
- assess* OR measure* OR tool OR outcome OR index OR test OR scale
AND
- stroke OR cerebrovascular * OR hemi* OR parkinson* OR sclerosis OR head injury OR brain injury OR motor neurone disease OR neuro*

The titles, then abstracts and full texts, were independently screened by the authors for articles that met the inclusion criteria. In addition, we searched the reference lists of previous reviews and the articles selected for the full text screening and searched the databases using the names of the selected measurement tools. If necessary, we contacted the original authors for clarification regarding the eligibility and for further data.

A description of the selected measurement tools, the participants and data about the psychometric properties and clinical utility were independently extracted by the authors from the selected articles. If the results of the selection, extraction or analysis differed, disagreements were resolved through discussion. A third person was available to arbitrate if agreement could not be reached.

First, we assessed the clinical utility of the measurement tools that met the selection criteria, to quantify the practicalities of using the measurement tools. Previously developed criteria based on the factors that influenced whether clinicians would use a measurement tool in clinical practice¹⁰ were used. These were as follows.

- Time taken to administer, analyse and interpret the measurement tool: 3 < 10 minutes; 2 = 10–30 minutes; 1 = 30–60 minutes; 0 > 1 hour.

- Cost: 3 < £100; 2 = £100–500; 1 = £500–1000; 0 > £1000.
- Does the measurement tool need specialist equipment and training to use? 2 = no; 1 = yes, but simple and clinically feasible; 0 = yes and not clinically feasible/unknown.
- Is the measurement tool portable? Can it be taken to the patient? 2 = yes easily (can fit into pocket); 1 = yes (in a briefcase or trolley); 0 = no or very difficult.
- Is the measurement tool accessible? Can detailed operating instructions for the measurement tools be obtained? 2 = yes (full standardized operating procedure/ instruction manual can be obtained from the article or a website); 1 = no, but operation can be simply worked out from a description in the article; 0 = no operating instructions available.

These scores were summated with a maximum of 12. Measurement tools that scores less than 10 were considered infeasible for use in clinical practice and were rejected at this stage. The remaining measurement tools were considered feasible and their psychometric properties were assessed to identify those which would provide robust data.

The strength of the psychometric properties were assessed using a previously described framework^{10,11} as follows:

- + weak reliability, validity or internal consistency = scores of < 0.4
- ++ moderate reliability, validity or internal consistency = scores of 0.4–0.6
- +++ good reliability, validity or internal consistency = scores of 0.6–0.8
- ++++ excellent reliability, validity or internal consistency = scores of > 0.8

As data from the tests of ability to detect change, content and construct validity and scaling properties are non-standardized, the acceptable (or unacceptable) limits were not specified but considered individually. The measurement properties were summarized to aid selection for use in practice. A measurement tool needed to obtain 'good' or 'excellent' scores for reliability and validity, and have some information about the ability to detect

Figure 1. Flowchart demonstrating the selection process.

change before it could be recommended for use in clinical practice.

Results

The searches revealed 25 articles that met the selection criteria, which assessed 17 measurement tools (summarized in Figure 1 and Table 1)^{12–36} including six versions of the Fatigue Severity Scale and five versions of the Fatigue Impact Scale. All involved patients' self-reported rating of their fatigue using a Likert scale delivered by

interview, post or telephone. They varied in length from five to 44 items. The aspects of fatigue measured were varied including the frequency, severity and variability of fatigue-related phenomena; impact of fatigue on everyday life; social consequences of fatigue; exacerbating and relieving factors; and agreement with experiences of fatigue. The timescale that patients were asked to recall ranged from a week, two weeks and a month. Most tools had been developed to evaluate fatigue in other conditions and applied to people with neurological conditions.

Table 1. Details of the selected measurement tools.

Measurement Tool	Construct measured	Content	Item generation	Scoring	Availability
Fatigue assessment scale ^{26,35,38}	Frequency of fatigue-related phenomena. Timescale unspecified	10 items regarding: activity; energy; bothersomeness; exhaustion; starting things; thinking motivation, concentration	Developed from an item pool of established measures of fatigue in the general population. No apparent patient involvement	5-point Likert scale where 0 = 'never' to 4 = 'always'. Individual scores (inappropriately) summated to give an overall score	Details available from Michielsen et al. ³⁹
Fatigue Impact Scale ^{21,25,33}	Impact of fatigue on everyday life in the previous month	44 items in three subscales; cognitive, physical and psychosocial	Developed from previously published fatigue scales plus interviews with people with MS	5-point Likert scale; 0 = 'no problem' to 4 = 'extreme problem'. Individual scores (inappropriately) summated to give an overall score	List of items and scoring scale available but not full operating instructions unavailable
Modified Fatigue Impact Scale ^{12,16,34}	Frequency of fatigue in everyday life in the previous month	21 items in three subscales; cognitive (10 items), physical (9 items) and psychosocial (2 items)	From the original FIS, number of items reduced and wording of the Likert scale changed from severity to frequency. No rationale for changes given	5-point Likert scale for individual items; 0 = 'never' to 4 = 'always'. Individual scores (inappropriately) summated to give an overall score	No details of operating instructions available in articles or details of how they could be obtained
Modified Fatigue Impact Scale (mFIS) (Phys-8 and Cog-5) ²⁹	Frequency of fatigue in everyday life in the previous month in people with multiple sclerosis	13 items in a physical scale (8 items) and a cognitive scale (5 items)	mFIS reduced following Rasch analysis	5-point Likert scale for individual items; 0 = 'never' to 4 = 'always'. Scores can be summated but table to convert ordinal to interval data is not available	No details of operating instructions available in articles or details of how they could be obtained
Modified Fatigue Impact Scale for Spinal Cord Injury (mFIS-SCI) [19]	Frequency with which fatigue has affected the patient in the past week	From the original mFIS (for multiple sclerosis). Still 21 items in three subscales (cognition; physical and psychosocial) but 3 items removed and 3 new items added. 2 items were re-worded. Wording of the Likert scale changed from 'severity' to 'frequency'. Time scale changed from four weeks to one	The mFIS- multiple sclerosis was revised for relevance for people with traumatic SCI by expert clinicians and patients	5-point Likert scale for individual items; 0 = 'never' to 4 = 'almost always'. Scores are (inappropriately) summated to produce an overall score	Details available in the article.
Unidimensional Fatigue Impact Scale (FIS) ^{27,36}	Frequency of fatigue-related phenomena. Timescale unspecified	22 items regarding: energy; pacing; activities; workload; cognition; change and emotions	Items generated from the original FIS plus patient interviews, and Rasch analysis of the resulting scale	4-point Likert scale; 0 = 'never' to 3 = 'all of the time'. The table to convert ordinal data to interval to enable summation available through the authors	Details from the authors
Fatigue Impact Scale for Daily Use (D-FIS) ^{14,24,26}	Severity of the impact of fatigue on daily life	8 items. Same items as the FIS but timescale and descriptors of the Likert scale changed	Items from the original FIS. No rationale for the choice of items	5-point Likert scale for individual items; 0 = 'no problem' to 4 = 'extreme problem')	No details of operating instructions available in articles or details of how they could be obtained

(Continued)

Table 1. (Continued)

Measurement Tool	Construct measured	Content	Item generation	Scoring	Availability
Fatigue Severity Scale ^{12,13,16,18,20,21}	Interferences with daily life. No timescale specified	9 items regarding motivation; physical function; responsibilities; work, family or social life; exercise; how easily fatigued; frequency of problems; priority of symptoms	Items selected for their common features between fatigue in multiple sclerosis and Systemic Lupus Erythematosus (SLE). No details regarding generation of items	7-point Likert scale; 1 = 'strongly disagree' to 7 = 'strongly agree'. Scores summated (inappropriately) to produce an overall score	List of the items and scoring scale available in the original article (Krupp et al. ²⁰) Full operating instructions unavailable
Fatigue Severity Scale (FSS)-5 ²⁸	Social consequences of fatigue. No timescale specified	5 items	Items of the original FSS reduced through Rasch analysis	Scoring unstated but assumed to be the same as the FSS. The table to convert ordinal data to interval to enable summation available in the article	A list of the items found in the article but not how they should be operationalized. Operating instructions not available elsewhere
Fatigue Severity Scale (FSS)-7 ²²	Social consequences of fatigue. No timescale specified	7 items	Items of the original FSS reduced through Rasch analysis	Scoring unstated but assumed to be the same as the FSS. The table to convert ordinal data to interval to enable summation available in the article	A list of the items included in the article but not how they should be operationalized. Operating instructions not available elsewhere
Multiple sclerosis specific Fatigue Severity Scale (FSS) ¹⁶	Exacerbating and relieving factors	6 items. Patients rate agreement with statements	Despite the name, no connection with the original FSS. No rationale given for choice of items	7-point Likert scale; 1 = 'strongly disagree' to 7 = 'strongly agree'. Scores summated (inappropriately) to produce an overall score	No details of operating instructions available in article or details of how they could be obtained
Neurological Fatigue Index for multiple sclerosis ^{30,31}	Experience of fatigue over the previous two weeks	5 subscales: physical (8 items); cognitive (4 items); diurnal rest (5 items) and abnormal sleep (5 items) and 10-item summary scale (physical and cognitive items). Patients rate agreement with statements	Developed from scratch with people with multiple sclerosis and expert clinicians	4 point Likert scale 0 = 'strongly disagree' to 3 = 'strongly agree'. Table to convert ordinal raw scores to interval so scores can be summated available in the article	Free for use in all public health and not-for-profit agencies. Available from the authors
Neurological Fatigue Index (NFI) for motor neurone disease (MND) ¹⁷	Experience of fatigue. Time scale not stated	3 subscales: weakness (7 items); energy (6 items); and summary scale (8 items). Patients rate their agreement with statements	Developed from scratch with people with MND	Scoring structure not stated. Assumed to be the same as other NFIs. Table to convert ordinal raw scores to interval so scores can be summated available in the article	Free for use in all public health and not-for-profit agencies. Available from the authors

Table 1. (Continued)

Measurement Tool	Construct measured	Content	Item generation	Scoring	Availability
Neurological Fatigue Index (NFI) for stroke ³²	Experience of fatigue over the previous 4 weeks	3 subscales: physical (8 items); cognitive (4 items); and 10-item summary scale). Patients rate their level of agreement with statements	Developed from the NFI-multiple sclerosis with stroke survivors	4 point Likert scale 0 = 'strongly disagree' to 3 = 'strongly agree'. Table to convert ordinal raw scores to interval so scores can be summated available in the article	Free for use in all public health and not-for-profit agencies. Available from the authors
Profile of Mood States (fatigue subscale) ²⁶	Assesses how much the person has had the feelings in the past week	6 items: tiredness; fatigue; worn-out; sluggish; weary; sleepy	Originally items generated by academics to measure of mood in people receiving psychotherapy. Fatigue is one of five subscales. No apparent patient involvement	4 point Likert scale. 0 = 'not at all'; 3 = 'extremely'. Individual scores (inappropriately) summated to give an overall score	List of the items and scoring scale available. ⁴⁰ Full operating instructions unavailable
Multidimensional Assessment of Fatigue (for Traumatic Brain Injury (TBI)) ²³	Frequency, severity and variability of fatigue; distress and interferes with daily life	15 items. Items are summated to produce a 'Global Fatigue Index'	Originally developed for people with cancer. No patient or expert clinician involvement	4 point Likert scale 0 = 'not at all' and 3 = 'a great deal'	No details of operating instructions available in article or details of how they could be obtained
Functional Assessment of Chronic Illness Therapy – Fatigue Scale (FACIT-F) ^{15,18}	Experience of fatigue and its impact on daily life	13 items. No further details available	Originally developed for people with cancer. No evidence of patient or clinicians involvement	5 point intensity rating (0–4). No further details	Details can be downloaded from http://www.fact.org/FACTOrg/Questionnaires

All the tools were simple 'paper and pencil' ordinal tests, which were freely available, and thus scored highly on the clinical utility criteria for cost, portability and specialist equipment. Time to complete the tests was not addressed, but from the details given one would expect them to be quick to perform, although the longer ones would probably take more than 10 minutes, especially for those with communication, cognitive, visual difficulties or upper limb weakness. Accessibility of the operating instructions for the tools was more limiting. Full operating instructions can be obtained from the authors for the Neurological Fatigue Indices^{17,30–32} and the Uni-dimensional Fatigue Impact Scale;^{27,36} from the web for Functional Assessment of Chronic Illness Therapy (Fatigue subscale)^{15,18} For the other scales, operating instructions were not available, although one could, to some extent, work out how the tools were intended to be used from their descriptions in the selected or preceding articles (detailed in Table 1).

The data extracted for the psychometric evaluation of the selected tools are summarized in Table 2 and detailed in Table 3 (available online). None of the tools met the criteria for all the psychometric properties. Older publications tended to use classic test theory to develop the measurement tools, focusing on the criterion-related and concurrent validity. The measures of fatigue tended to show good associations with each other (criterion-related validity); moderate–good associations with measures of mood and quality of life and weaker associations with measures of disability and motor impairment (concurrent validity).

More recent publications have used Item Response Theory, specifically Rasch analysis to develop and evaluate content and scale structure. This revealed that well-established tools, most notably the Fatigue Severity Scale^{12,28} and the Fatigue Impact Scale^{27,29,34} do not fit the Rasch model. Several authors have attempted to rectify the situation by adapting these tools, but none have emerged as a definitive version.^{12–14,19,24,27–29,34} All the tools summate the scores from the items to produce a total score; but only the Neurological Fatigue Indices,^{17,30–32} Uni-dimensional Fatigue Impact Scale^{27,36} and the 5- and 7-item versions of

the Fatigue Severity Scale,^{22,28} included tables to enable the ordinal data from the items to be converted to interval level data, which is needed statistically before summation and parametric analysis.³⁷

Reliability has been less thoroughly evaluated than validity, and inter-tester reliability less than test–retest. However, when reported, both were good–excellent (Table 2). Ability to detect change has seldom been addressed, but when calculated, it appears that relatively low changes in score are needed to overcome the measurement error. However, there are quite widely differing estimates in different populations for the same scale. The clinical or personal significance of any changes has not been addressed.

Discussion

This systematic review has revealed many tools to measure fatigue in neurological conditions, most of which were modifications of the Fatigue Severity and Fatigue Impact Scales. All the tools, apart from the Neurological Fatigue Indices,^{17,30–32} were originally developed to evaluate fatigue in other conditions and then applied to people with neurological conditions. In doing so, scant consideration was given to the construct or patients' experience of fatigue in neurological conditions; it was assumed to be the same as for non-neurological conditions. Consequently, many items refer to the impact of fatigue on physical, mental and emotional function, and everyday activities and participation, which are also often directly limited by neurological conditions rather than (or in addition to) fatigue, thus conflating the fatigue scores.

Early work using classic test theory appeared to support this transfer and the two main tools; the Fatigue Severity Scale^{16,18,20,21} and Fatigue Impact Scale^{21,25,33} became widely used. However, the advent of Item Response Theory has highlighted their shortcomings, including redundant items, disordered scoring structures and an inadequate underlying construct.^{12,19,27–29} Many authors have attempted to rectify the situation by adapting the tools to the Rasch model and/or for specific clinical groups, which has produced many versions with

Table 2. Summary of the psychometric properties of the selected measurement tools. Full details are found in Table 3, available online.³⁸

Scale	Validated for	Criterion-related validity	Concurrent validity	Internal consistency	Test-retest	Inter-tester	Responsiveness	Scaling structure
Fatigue Assessment Scale ^{26,35,38}	Stroke	Good	QoL: moderate Mood: moderate	Moderate-good	Weak-good	Good-excellent	MDD = 14 points	Multi-dimensional (4 items relate to mood)
Fatigue Impact Scale ^{21,25,33}	MS	Weak-good	Motor activity: weak QoL: moderate		Good-excellent			
Modified Fatigue Impact Scale ^{21,16,34}	MS PD	Weak-good Moderate	Pain: moderate Mood: weak-good Disability: weak Mood: good	Excellent Good				No floor and ceiling effects Not Uni-dimensional (two factors) Not Uni-dimensional (two factors) Final version fits the Rasch model
Modified Fatigue Impact Scale (Physical-8 and Cognitive-5) ²⁹	MS							
SCI version of the mFIS ¹⁹	SCI injuries	Moderate-excellent	Mood: weak	Excellent				Strong floor effect
Unidimensional Fatigue Impact Scale ^{27,36}	MS		QoL: weak-good Mood: moderate-good Disability: weak-good	Excellent			MDD = 7 points	Final version fits the Rasch model
Fatigue Impact Scale for Daily Use ^{14,24,26}	MS PD	Moderate-good Moderate	Mood: moderate QoL: good Disability: weak Disability: weak Mood: weak	Excellent Excellent	Good-excellent		MDD and SEM = 3.2 points SEM = 2.15 points	No floor and ceiling effects No floor and ceiling effects No redundancy No floor and ceiling effects Unidimensional No redundancy
Fatigue Severity Scale ^{12,13,16,18,20,21}	MS SCI ABI PD	Weak-good Good Weak-good Good	Pain: moderate Mood: weak-moderate Disability: weak Mood: moderate Motor activity: weak	Excellent Excellent Excellent	Excellent		MDD = 2 points	No floor and ceiling effects but imprecise for very severe fatigue Does not fit the Rasch model as measures several constructs Fits the Rasch model. No floor or ceiling effects Fits the Rasch model but has a large ceiling effect Fits the Rasch model
Fatigue Severity Scale-5 ²⁸	MS							
Fatigue Severity Scale-7 ²²	Stroke		QoL: moderate					
MS Specific Fatigue Severity Scale ¹⁶	MS	Very weak						
Functional Assessment of Chronic Illness-Fatigue ^{15,18}	Stroke PD	Good	Global function: moderate Global function: good	Excellent Excellent	Excellent			Fits the Rasch model No floor or ceiling effects

(Continued)

Table 2. (Continued)

Scale	Validated for	Criterion-related validity	Concurrent validity	Internal consistency	Test–retest	Inter-tester	Responsiveness	Scaling structure
Multidimensional Assessment of Fatigue Scale ²³	ABI			Excellent				No floor or ceiling effects Final scale fits the Rasch model
Neurological Fatigue Index MND ¹⁷	MND	Moderate–good	Disability: weak	Excellent			MDD 1 point	Fits the Rasch model
Neurological Fatigue Index-MS ^{30,31}	MS	Moderate					MDD = 2.5 points	Fits the Rasch model
Neurological Fatigue Index-Stroke ³²	Stroke	Weak–good	QoL: moderate–good					Fits the Rasch model
Profile of Mood States (fatigue subscale) ²⁶	Stroke	Moderate–good	Moderate	Moderate–good	excellent			

ABI: acquired brain injury; mFIS: modified Fatigue Impact Scale; MDD: minimum detectable difference; MND: motor neurone disease; MS: multiple sclerosis; PD: Parkinson's disease; QoL: quality of life; SEM: standard error of measurement; SCl: spinal cord injury.

confusingly similar names and abbreviations, but none have emerged as a definitive version and none are recommended.

All the scales summate the item scores to produce a total score. This is a controversial method that many would consider an inappropriate use of data, as the calculation assumes continuous data, while the item scores are categorical.³⁷ Only the Neurological Fatigue Indices,^{17,30–32} Uni-dimensional Fatigue Impact Scale^{27,36} and the 5- and 7-item versions of the Fatigue Severity Scale^{22,28} have published data to enable the ordinal item scores to be converted to interval level so that item scores can be appropriately summated and parametric statistics used. For the other scales, the total items scores should be used with caution until the effect, if any, of using ordinal data as if it were continuous is established. If such use of data proves inappropriate, it could lead to ineffective clinical decisions or inaccurate research conclusions being made. In addition, future publications should require inclusion of the data conversion tables.

None of the selected tools met all the criteria demonstrating robust psychometrics and utility. However, the Neurological Fatigue Indices were the most thoroughly developed, and could be cautiously used for people with multiple sclerosis, stroke and motor neurone disease.^{17,30–32} They were developed with patients, demonstrate a robust scale structure, produce interval data and are valid, easy to use and freely available from the authors. The authors reported excellent test–retest reliability, but used measures of association^{17,30–32} rather than agreement; the *a priori* definition used in this review. Further research is needed to demonstrate the reliability (in terms of agreement) and the inter-rater reliability. Nevertheless, the methodology used for the Neurological Fatigue Indices are an exemplar for the development of patient-reported tools for other conditions.

Inter-rater reliability was the least frequently evaluated psychometric property. As fatigue is self-reported, one could argue that inter-tester reliability is not an issue. However, as many neurological patients struggle to complete tools, others often assist them, and the way in which they do so is likely to impact on the scores obtained. Further

research is needed to address the inter-rater reliability and the impact of mode of delivery (by interview, post, telephone or proxy, for example). Future publications should require full operating instructions to be available to facilitate consistent delivery.

Given the inherent subjectivity of fatigue, it is fundamental that tools should reflect patients' experience, and change in score should reflect meaningful change for the patient. The importance of the patients' perspective is a relatively recent phenomenon in tool development and this is reflected in the selected articles. It was rarely part of the older tools, except the Fatigue Impact Scale, which evolved, in part, from interviews with patients with multiple sclerosis.³⁸ However, more recent articles, most notably the Neurological Fatigue Indices,^{17,30–32} included patients from the outset to inform the tools' content.

Robust patient-based fatigue measurement tools are needed for all neurological conditions, as it appears to be an almost universal feature. This raises the intriguing question whether this common phenomenon has a common mechanism, for which a single universal measure should, or could, be developed. The Neurological Fatigue Indices for stroke and multiple sclerosis^{30–32} have common items, while variations of the Fatigue Impact Scale and Fatigue Severity Scale are valid for people with multiple sclerosis, spinal cord injury, acquired brain injury and Parkinson's Disease,^{12–14,16,18,20,21,24} suggesting a common construct. In contrast, the Neurological Fatigue Index for Motor Neurone Disease¹⁷ has a different structure, and variations of the Fatigue Impact^{12,19,29,34} and Fatigue Severity Scales^{16,22,28} for specific conditions have produced different contents and structures, which refutes the notion of a commonality. Further research to explore patients' experience of fatigue, their needs and treatment priorities and the mechanisms of fatigue in different neurological conditions is needed. In the meantime disease specific tools should be used; it cannot be assumed that a fatigue measure that is effective in one condition would be suitable for use with another.

Several limitations need to be borne in mind when considering these results. First, we only included

publications in English and also did not attempt to identify unpublished data or explore the grey literature, so although we had thorough search strategies we may have missed relevant publications in other languages or have a publication bias in the identified data. Second, our quality judgments of the psychometric properties and clinical utility were arbitrary. Although they have strong face validity, we do not know whether our choices are optimal; different choice could have produced different conclusions.

Clinical messages

- A total of 17 tools assessing fatigue in neurological conditions were identified, including six versions of the Fatigue Severity Scale and five versions of the Fatigue Impact Scale.
- None of the selected tools met all of the psychometric and utility criteria. The Neurological Fatigue Indices were the best available.

Conflict of interest

The author declares that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Chaudhuri A and Behan OP. Fatigue in neurologic disorders. *Lancet* 2004; 363: 978–988.
2. Intercollegiate Stroke Working Party (2012). *National clinical guideline for stroke*. 4th ed. London: Royal College of Physicians.
3. Turner-Stokes L, Nair A, Sedki I, et al. Multi-disciplinary rehabilitation for acquired brain injury in adults of working age. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD004170.
4. Burton L, Tyson SF and McGovern A. Staff perceptions of using outcome measures in stroke rehabilitation. *Disab Rehab* 2013; 35(1): 828–34. DOI: 10.3109/09638288.2012.709305.
5. Dittner A, Wessely S and Brown R. The assessment of fatigue: A practical guide for clinicians and researchers. *J Psychosom Res* 2004; 56(2) 157–170.
6. Friedman J, Alves G, Hagell P, et al. Fatigue rating scales critique and recommendations by the movement disorders society task force on rating scales for Parkinson's disease. *Mvt Dis* 2010; 25(7): 805–822.
7. Kos D, Kerckhofs E, Ketelaer P, et al. Self-report assessment of fatigue in multiple sclerosis: A critical evaluation. *Occ Ther Health Care* 2004; 17(3–4): 45–62.
8. Elbers RG, Rietberg MB, van Wegen EHE, et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Qual Life Res* 2012; 21: 925–944.
9. Terwee C, Mokkink L, Knol D, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21(4): 651–657.
10. Tyson S, Watson A, Moss S, et al. Development of an evidence based framework for the physiotherapy assessment of neurological conditions? *Disab Rehabil* 2007; 30(2): 142–144.
11. Connell L and Tyson SF. Measures of sensation in neurological conditions: a systematic review. *Clin Rehab* 2012; 26(1): 68–80.
12. Antmann D, Bamer AM, Noonan V, et al. Comparison of the psychometric properties of two fatigue scales in multiple sclerosis. *Rehab Psych* 2012; 57(2): 159–166.
13. Anton HA, Miller WC and Townson AF. Measuring fatigue in persons with spinal cord injury. *Arch Phys Med Rehabil* 2008; 89(3): 538–542.
14. Benito-León J, Martínez-Martin P, Frades B, et al. Impact of fatigue in multiple sclerosis: The Fatigue Impact Scale for Daily Use (D-FIS). *Mult Scler* 2007; 13: 645–651.
15. Butt Z, Lei JS, Rao D, Heinemann AW, et al. Measurement of fatigue in cancer, stroke and HIV using the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F) scale. *J Psychos Res* 2013; 74: 64–68.
16. Flachenecker P, Kämpfel T, Kallmann B, et al. Fatigue in multiple sclerosis; a comparison of different rating scales and correlation to clinical parameters. *Mult Scler* 2002; 8: 523–526.
17. Gibbons CJ, Mills RJ, Thornton EV, et al. Development of a patient reported outcome measure for fatigue in motor neurone disease: the Neurological Fatigue Index (NFI-MND). *Health Qual Life Outc* 2011; 9: 101.
18. Hagell P, Hoglund A, Reimer J, et al. Measuring fatigue in Parkinson's Disease: A psychometric study of two brief generic fatigue questionnaires. *J Pain Sympt Manag* 2006; 32(5): 420–432.
19. Imam B, Anton HA and Miller WC. Measurement properties of a telephone version of the modified Fatigue Impact Scale among individuals with a traumatic spinal cord injury. *Spin Cord* 2012; 50: 920–924.
20. Krupp LB, LaRocca NG, Muir-Nash J, et al. The Fatigue Severity Scale. *Arch Neurol* 1989; 46: 1121–1123.
21. LaChapelle DL and Finlayson MAJ. An evaluation of subjective and objective measures of fatigue in patients with brain injury and healthy controls. *Brain Inj* 1998; 12(8): 649–659.

22. Lerdal A and Kottorp A. Psychometric properties of the Fatigue Severity Scale – Rasch Analyses in a Norwegian Stroke Cohort. *International J Nurs Stud* 2011; 48: 1258–1265.
23. Lequerica A, Bushnik T, Wright J, et al. Psychometric Properties of the Multidimensional Assessment of Fatigue Scale in Traumatic Brain Injury. *J Head Trauma Rehabil* 2012; 27(6): E28–E35.
24. Martinez-Martin P, Catalan MJ, Benito-Leon J, et al. Impact of fatigue in Parkinson's Disease: The Fatigue Impact Scale for Daily Use (D-FIS). *Qual Life Res* 2006; 15: 597–606.
25. Mathiowetz V. Test–retest reliability and convergent validity of the Fatigue Impact Scale for persons with multiple sclerosis. *Am J Occ Ther* 2003; 57(4): 389–395.
26. Mead G, Lynch J, Greig C, et al. Evaluation of fatigue scales in stroke. *Stroke* 2007; 38(7): 2090–2095.
27. Meads DM, Doward LC, McKenna SP, et al. The development and validation of the Uni-dimensional Fatigue Impact Scale (U-FIS). *Mult Scler* 2009; 15(10): 1228–1238.
28. Mills RJ, Young CA, Nicholas RS, et al. Rasch analysis of the Fatigue Severity Scale in multiple sclerosis. *Mult Scler* 2009; 15: 81–87.
29. Mills RJ, Young CA, Pallant JF, et al. Rasch analysis of the Modified Fatigue Impact Scale (mFIS) in multiple sclerosis. *J Neurol Neurosurg Psych* 2010; 81(9): 1049–1051.
30. Mills RJ, Young CA, Pallant JF, et al. Development of a patient reported outcome measure scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). *Health Qual Life Outc* 2010; 8: 22.
31. Mills RJ, Calabresi M, Tennant A, et al. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index. *Mult Scler* 2012; 18: 862–870.
32. Mills RJ, Pallant JF, Koufali M, et al. Validation of the Neurological Fatigue Index for stroke (NFI-Stroke). *Health & Qual Life Outc* 2012; 10: 51.
33. Rendas-Baum R, Yang M, Cattelin F, et al. A novel approach to estimate the minimally important difference for the Fatigue Impact Scale in multiple sclerosis. *Qual Life Res* 2010; 19: 1349–1358.
34. Schiehser DM, Ayers CR, Lui L, et al. Validation of the Modified Fatigue Impact Scale in Parkinson's disease. *Parkinson Rel Dis* 2013; 19(3): 335–338.
35. Smith ORF, van den Broek KC, Renkens M, et al. Comparison of fatigue levels in patients with stroke and patients with end stage heart failure: application of the Fatigue Assessment Scale. *JAGS* 2008; 56: 1915–1919.
36. Twiss J, Doward LC, McKenna SP, et al. Interpreting scores on multiple sclerosis-specific patient reported outcome measures (the PRIMUS and U-FIS). *Health & Qual Life Outc* 2010; 8: 117–124.
37. Tennant A and Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007; 57(8): 1358–1362.
38. Fisk JD, Ritvo PG, Ross L, et al. Measuring the functional impact of fatigue: initial validation of the Fatigue Impact Scale. *Clin Infect Dis* 1994; 18: S79–83.
39. Michielsen HJ, De Vries J and Van Heck GL. Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *J Psychosom Res* 2003; 54(4): 345–352.
40. McNair DM and Lorr M. An analysis of mood in neurotics. *J Abnormal Social Psychol* 1964; 69: 620–627.