arm

Annals of Rehabilitation Medicine

Psychometric Properties of Three Fatigue Rating Scales in Individuals With Late Effects of Polio

Jan Lexell, MD, PhD[1,2,3,4], Stina B. Jonasson, RPT, PhD[1,4], Christina Brogardh, RPT, PhD[1,4]

[1]Department of Health Sciences, Lund University, Lund; [2]Department of Neuroscience, Rehabilitation Medicine, Uppsala University, Uppsala; [3]Department of Health Science, Lulea University of Technology, Lulea; [4]Department of Neurology and Rehabilitation Medicine, Skane University Hospital, Lund, Sweden

Objective To evaluate the psychometric properties of the Fatigue Severity Scale (FSS), the Fatigue Impact Scale (FIS), and the Multidimensional Fatigue Inventory (MFI-20) in persons with late effects of polio (LEoP). More specifically, we explored the data completeness, scaling assumptions, targeting, reliability, and convergent validity.

Methods A postal survey including FSS, FIS, and MFI-20 was administered to 77 persons with LEoP. Responders received a second survey after 3 weeks to enable test-retest reliability analyses.

Results Sixty-one persons (mean age, 68 years; 54% women) responded to the survey (response rate 79%). Data quality of the rating scales was high (with 0%–0.5% missing item responses), the corrected item-total correlations exceeded 0.4 and the scales showed very little floor or ceiling effects (0%–6.6%). All scales had an acceptable reliability (Cronbach's $\alpha \geq 0.95$) and test-retest reliability (intraclass correlation coefficient, ≥ 0.80). The standard error of measurement and the smallest detectable difference were 7%–10% and 20%–28% of the possible scoring range. All three scales were highly correlated (Spearman's correlation coefficient $r_s = 0.79$–0.80; $p < 0.001$).

Conclusion The FSS, FIS, and MFI-20 exhibit sound psychometric properties in terms of data completeness, scaling assumptions, targeting, reliability, and convergent validity, suggesting that these three rating scales can be used to assess fatigue in persons with LEoP. As FSS has fewer items and therefore is less time consuming it may be the preferred scale. However, the choice of scale depends on the research question and the study design.

Keywords Fatigue, Postpoliomyelitis syndrome, Psychometrics, Rehabilitation, Reliability of results

INTRODUCTION

Several years after an acute poliomyelitis infection, pa-

tients may experience new symptoms such as pain, muscle weakness, muscle fatigue as well as general fatigue, which are commonly referred to as late effects of polio

(LEoP) or postpolio syndrome [1]. Among these symptoms, fatigue is often reported as the most disabling [2-6] and chronic challenge [7]. Fatigue has been defined as "an overwhelming sense of tiredness, lack of energy and feeling of exhaustion" [8]. As fatigue is negatively associated with mobility [4], quality of life [9,10] and life satisfaction [5], it is important to evaluate fatigue and to plan appropriate interventions that reduce its impact.

As fatigue is mainly a subjective experience, self-report rating scales are used to assess fatigue. The three most commonly used scales to assess fatigue in persons with LEoP include the Fatigue Severity Scale (FSS) [11], the Fatigue Impact Scale (FIS) [12], and the Multidimensional Fatigue Inventory (MFI-20) [13]. To facilitate accurate assessments, the self-report rating scales need to be psychometrically sound. The validity and reliability of FSS, FIS, and MFI-20 have been studied in persons with LEoP [14-19], but comprehensive analyses of other psychometric properties of the scales are unavailable. Factors such as the number of missing items, the score distribution, any skewness, and floor and ceiling effects are important to evaluate. Moreover, these scales are considered for the assessment of similar underlying construct, i.e., fatigue. However, no study has explored their convergent validity in terms of the relationship between the scales. To enhance our understanding and support the choice of scales in clinical research, further evaluations of their psychometric properties are required.

The aim of this study was to evaluate the psychometric properties of the FSS, FIS, and MFI-20 in persons with LEoP. More specifically, we explored the data completeness, scaling assumptions, targeting, reliability, and convergent validity. Our hypothesis was that all scales are psychometrically sound and have high convergent validity.

MATERIALS AND METHODS

The study was approved by the Regional Ethical Review Board of Lund University, Sweden (No. Dnr 2013-380). All participants provided written informed consent.

Participants

Potential participants were recruited from a clinical database at a postpolio clinic in a university hospital in southern Sweden. The inclusion criteria were as follows:

a confirmed history of acute poliomyelitis, a period of recovery and functional stability of at least 15 years, clinically verified LEoP with new symptoms that persisted for at least 1 year, and age between 50 and 80 years. The exclusion criteria were as follows: other major diseases that lead to severe fatigue, known cognitive dysfunction, and difficulties in reading and writing Swedish. One of the authors (JL, physician with knowledge of the subjects' medical history) screened the database for potential participants together with the team physiotherapist. In total, 232 persons met the criteria and every third person among them (i.e., n=77) was invited to participate in the study. Based on our previous studies in people diagnosed with LEoP, we anticipated a response rate of over 70%. Thereby, we were able to recruit at least 50 participants with total scores on all the three scales needed for the analyses of the psychometric properties [20] and 30 participants needed for the test-retest reliability analysis [21].

Procedure

Data were collected through a postal survey. The 77 potential participants were mailed the following details: study information, an informed consent form, socio-demographic and disease-related questions, the three fatigue-rating scales (FSS, FIS, and MFI-20) and a pre-stamped envelope to return the questionnaires and fatigue rating scales; and is hereafter referred to as t1. After 3 weeks, all the responders at t1 received a second survey containing only the fatigue rating scales; this is hereafter referred to as t2. A reminder was sent after 2 weeks to non-responders at both t1 and t2.

Socio-demographic and disease-related questions

Socio-demographic questions targeted marital status, vocational situation, diseases other than LEoP, current medication, use of nighttime respiratory ventilator, use of orthotics and mobility devices indoors and/or outdoors, and walking ability. The Self-reported Impairments in Persons with late effects of Polio (SIPP) rating scale was used to assess the extent to which the participants were bothered by various LEoP-related impairments [22].

The three fatigue rating scales
The Fatigue Severity Scale
FSS consists of 9 statements (items), e.g., "I am easily fatigued" and "Fatigue interferes with my work, family,

or social life". Items are scored on a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The total score represents the mean of the nine statements and ranges from 1 to 7, where a greater score indicates more fatigue [11]. The Swedish translated FSS was used in this study [23].

The Fatigue Impact Scale

FIS consists of 40 statements (items), ten each in the subscales that cover cognitive and physical dimensions, respectively, and 20 in the subscale that covers a social dimension. Items include, e.g., "Because of my fatigue, I feel less alert" (cognitive dimension), "Because of my fatigue, I have to limit my physical activities" (physical dimension), and "Because of my fatigue, minor difficulties seem like major difficulties" (social dimension). Possible response options range from 0 (no problem) to 4 (extreme problem). Items are summed into a total score ranging from 0 to 160, where a greater score indicates more fatigue. The cumulative score of the Swedish translated FIS was used in this study [24].

The Multidimensional Fatigue Inventory

MFI-20 consists of 20 statements (items), four in each of the 5 subscales that cover general, physical, and mental fatigue, reduced activity, and reduced motivation, respectively. Possible response options range from 1 (yes, that is true) to 5 (no, that is not true). The scale includes an equal number of items that are indicative of fatigue (e.g., "I tire easily" and "I don't feel like doing anything") and contra-indicative (e.g., "I feel very active" and "I can concentrate well"), respectively [13,25]. Items that are indicative of fatigue (i.e., items #2, 5, 9, 10, 13, 14, 16-19) are recoded so that the response option "no, that is not true" equals 1, and "yes that is true" equals 5. Items for each subscale are summed into subscale scores ranging from 4 to 20, where a greater score indicates more fatigue. A recent Rasch analysis of the scale has shown that MFI-20 can be considered unidimensional, suggesting that raw scores can be transformed into interval scores, and the total (transformed) cumulative score can be used as a global measure of fatigue [19]. The total score ranges from 20 to 100, where a greater score indicates more fatigue. The Swedish translated MFI-20 was used in this study [26].

Analyses

Data completeness, scaling assumptions, targeting, internal consistency, reliability, and convergent validity were analyzed using data from t1. Data from both t1 and t2 were used for the analyses of test-retest reliability. Transformation of the total scale scores for MFI-20 into intervals was used in the analyses of total mean score, min-max, skewness, and test-retest reliability, whereas raw scores were used in the analyses that involved item scores. All analyses were performed using the SPSS Statistics version 23 (IBM, Armonk, NY, USA). The level of statistical significance was set to $p<0.05$.

Data completeness

Data completeness was calculated as the percentage of missing item responses and the percentage of participants who obtained total scores [27,28]. Imputations were not used, i.e., a total score requiring response to every item.

Scaling assumptions

Requirements for the legitimacy of cumulative total score were explored. Item means and standard deviations (SD) were roughly parallel within a scale. Further, items contribute adequately to the total score and measure the same underlying construct. All these assumptions were fulfilled if corrected item-total correlations exceed 0.4 [27,28].

Targeting

Targeting refers to the ability of a scale's score distribution to reflect the true value, e.g., fatigue in a sample of patients [27]. Targeting was explored by studying the scales' score distribution, skewness, floor and ceiling effects. The total score spanned the full possible scoring range, mean scores were close to the midpoints, skewness was less than ±1 [27,29], and the floor and ceiling effects were less than 20% [27].

Reliability

Reliability was explored by studying the scales' internal consistency (assessed with Cronbach's α) [30] and test-retest reliability (assessed with one-way random, single measurement intraclass correlation coefficient [ICC]) [31]. Cronbach's α and ICC values above 0.70 are considered acceptable for group comparisons, while ICC

values above 0.90–0.95 are suggested as a minimum for a rating scale of individual comparisons [32]. Standard error of measurement (SEM) was calculated using the formula $SD_{baseline} \times \sqrt{1-reliability}$ [33] and smallest detectable difference (SDD) was calculated using the formula $SEM \times 1.96 \times \sqrt{2}$ [34]. SEM and SDD values were also expressed as a percentage of the possible scoring range in order to equalize differences based on the different scoring ranges in the scales, to ensure fair comparison. The scores were calculated as *SEM or SDD / number of possible scoring options* ×100. The mean differences (đ) in scale scores between t1 and t2 and the 95% confidence interval (CI) around đ were calculated to explore any systematic differences between the two test occasions; CI including 0 implies the absence of systematic differences [21].

Convergent validity

The convergent validity was assessed by determining the Spearman's correlation coefficient (r_s) between the three fatigue rating scales. The following limits were used for interpretation of the correlation coefficients: 0.0–0.3, negligible correlation; 0.3–0.5, low; 0.5–0.7, moderate; 0.7–0.9, high; and 0.9–1.0, very high correlation [35].

RESULTS

Of the 77 potential participants who received the first postal survey (t1), 14 did not respond and 2 explicitly declined to participate. Thus, a total of 61 persons (54% women; age, 68±5 years; range, 55–75 years) responded at t1 (response rate 79%). Their age at the acute poliomy-elitis infection was 5±3 years (min–max, 1–14 years), and the number of years before the onset of LEoP was 44±8 years (min–max, 30–60 years).

A majority used lower limb orthotics (61%) and outdoor mobility devices (52%), and 75% walk more than 100 m. Their SIPP score was 26±7 (min–max, 15–46), which indicates that they were moderately bothered by various LEoP-related impairments. The majority of the participants (n=46; 75%) reported comorbidities, e.g., cardiovascular disorders (n=31), diabetes (n=9), gastrointestinal disorders (n=7) and sleep apnea (n=7, including 5 using a night-time ventilator). Most participants (n=49; 80%) were treated with medication, mostly for hypertension (n=31), musculoskeletal pain (n=20), sleep disturbances (n=12), depression (n=4) and thyroid disease (n=4).

Of the 61 subjects that responded to the survey at t1, 56 responded to the second survey (t2) and thereby constitute the sample for the test-retest reliability analysis.

Data completeness

Data completeness for FSS was 100%, suggesting no missing item responses and all the 61 participants reported a FSS total score. The rate of missing item responses for FIS was 0.5%. A FIS total score was reported in 51 participants. The MFI-20 showed a 0.4% missing item response rate, while 58 participants obtained a total score. Data completeness of the three fatigue rating scales is presented in Tables 1–3.

Scaling assumptions

All three fatigue rating scales showed corrected item-total correlations exceeding 0.40; item means and SDs for

Table 1. Scoring distribution and data completeness of the Fatigue Severity Scale (n=61)

Item	Statement	Score	Missing response
1	My motivation is lower when I am fatigued	5.1±1.7	-
2	Exercise brings on my fatigue	3.9±2.0	-
3	I am easily fatigued	4.6±1.9	-
4	Fatigue interferes with my physical functioning	4.6±1.9	-
5	Fatigue causes frequent problems for me	4.1±1.9	-
6	My fatigue prevents sustained physical functioning	4.5±2.1	-
7	Fatigue interferes with carrying out certain duties and responsibilities	3.8±2.1	-
8	Fatigue is among my three most disabling symptoms	4.0±2.3	-
9	Fatigue interferes with my work, family, or social life	3.7±2.2	-

Values are presented as mean±standard deviation.
Possible item scoring range 1–7. Greater scores indicate more fatigue. Average item score is 4.3±2.0.

Table 2. Scoring distribution and data completeness of the Fatigue Impact Scale (n=61)

Item	Statement	Score	Missing response
	Because of my fatigue:		
1[a]	I feel less alert	1.9±1.0	-
2[c]	I feel that I am more isolated from social contact	1.4±1.1	-
3[c]	I have to reduce my workload or responsibilities	1.8±1.0[c]	-
4[c]	I am more moody	1.2±1.1	-
5[a]	I have difficulty paying attention for a long period of time	1.4±1.1	-
6[a]	I feel like I cannot think clearly	1.1±1.0	-
7[c]	I work less effectively (inside or outside the home)	1.6±1.0	-
8[c]	I have to rely more on others to help me or do things for me	1.4±1.1	1
9[c]	I have difficulty planning activities ahead of time because my fatigue may interfere with them	1.1±1.2	-
10[b]	I am more clumsy and uncoordinated	1.6±1.1	-
11[a]	I find that I am more forgetful	1.3±1.1[a]	-
12[c]	I am more irritable and more easily angered	1.0±1.1	-
13[b]	I have to be careful about pacing my physical activities	2.0±1.1	-
14[b]	I am less motivated to do anything that requires physical effort	2.0±1.1	-
15[c]	I am less motivated to engage in social activities	1.4±1.2	-
16[c]	My ability to travel outside my home is limited	1.6±1.3	-
17[b]	I have trouble maintaining physical effort for long periods	2.3±1.2	-
18[a]	I find it difficult to make decisions	1.0±1.0	-
19[c]	I have few social contacts outside of my own home	1.2±1.3	-
20[c]	Normal day-to-day events are stressful for me	1.2±1.1	-
21[a]	I am less motivated to do anything that requires thinking	1.0±1.0	-
22[c]	I avoid situations that are stressful for me	1.4±1.0	-
23[b]	My muscles feel much weaker than they should	2.0±1.2	1
24[b]	My physical discomfort is increased	1.7±1.2	2
25[c]	I have difficulty dealing with anything new	1.2±1.2[c]	-
26[a]	I am less able to finish tasks that require thinking	1.0±1.1	1
27[c]	I feel unable to meet the demands that people place on me	1.2±1.1	-
28[c]	I feel less able to provide financial support for myself and my family	0.9±1.2	1
29[c]	I engage in less sexual activity	1.6±1.3	1
30[a]	I find it difficult to organize my thoughts when I am doing things at home or at work	1.0±1.0	-
31[b]	I am less able to complete tasks that require physical effort	1.9±1.1	-
32[b]	I worry about how I look to other people	0.8±1.1	-
33[c]	I am less able to deal with emotional issues	0.9±1.1	-
34[a]	I feel slowed down in my thinking	1.1±1.1	-
35[a]	I find it hard to concentrate	1.2±1.1	-
36[c]	I have difficulty participating fully in family activities	1.3±1.2	2
37[b]	I have to limit my physical activities	2.2±1.1	1
38[b]	I require more frequent or longer periods of rest	1.8±1.1	-

Table 2. Continued

Item	Statement	Score	Missing response
39[c]	I am not able to provide as much emotional support to my family as I should	0.9±1.2	1
40[c]	Minor difficulties seem like major difficulties	1.2±1.1	1

Values are presented as mean±standard deviation.
Possible item scoring range 0–4. Greater scores indicate more fatigue. Average item score is 1.4±1.1. The missing item responses were spread among 10 participants.
[a]Cognitive dimension, [b]physical dimension, [c]social dimension.

Table 3. Scoring distribution and data completeness of the Multidimensional Fatigue Inventory (n=61)

Item	Statement	Score	Missing response
1[a]	I feel fit	3.3±1.4	1
2[b]	Physically I feel only able to do a little	2.9±1.4	-
3[c]	I feel very active	3.3±1.2	1
4[d]	I feel like doing all sorts of nice things	2.3±1.3	-
5[a]	I feel tired	3.5±1.4	1
6[c]	I think I do a lot in a day	3.5±1.3	-
7[e]	When I am doing something, I can keep my thoughs on it	2.1±1.2	-
8[b]	Physically I can take on a lot	3.9±1.1	-
9[d]	I dread having to do things	2.1±1.2	-
10[c]	I think I do very little in a day	2.9±1.4	-
11[e]	I can concentrate well	2.7±1.3	-
12[a]	I am rested	3.5±1.3	-
13[e]	It takes a lot of effort to concentrate on things	2.8±1.3	-
14[b]	Physically I feel I am in a bad condition	3.4±1.3	2
15[d]	I have a lot of plans	2.5±1.3	-
16[a]	I tire easily	3.8±1.3	-
17[c]	I get little done	3.0±1.4	-
18[d]	I don't feel like doing anything	2.3±1.4	-
19[e]	My thoughts easily wander	3.1±1.4	-
20[b]	Physically I feel I am in an excellent condition	3.8±1.2	-

Values are presented as mean±standard deviation.
Possible item scoring range 1–5. Greater scores indicate more fatigue. Contra-indicative items (#2, 5, 9, 10, 13, 14, 16-19) are recoded so that greater scores indicate more fatigue. Average item mean score is 3.0±1.3. The missing item responses were spread among three participants.
[a]General fatigue, [b]physical fatigue, [c]reduced activity, [d]reduced motivation, [e]mental fatigue.

the three scales are presented in Tables 1–3.

The mean scores and SDs for FSS were roughly parallel for all items; the mean scores ranged from 3.7 to 5.1 and SDs ranged from 1.7 to 2.3.

The SDs for FIS remained roughly parallel for all items, whereas the mean scores varied more across items. SDs ranged from 1.0 to 1.3 and mean scores ranged from 0.8 to 2.3. Twelve items (no. 6, 9, 12, 18, 21, 26, 28, 30, 32–34, and 39) included more respondents who selected the lower response options indicating less fatigue, resulting in a mean score that was >20% lower than the average item mean score (i.e., 1.4). Ten items (#1, 3, 13, 14, 17, 23, 24, 31, 37, and 38) included more respondents who selected the higher response options indicating more

fatigue, resulting in a mean score that was >20% higher than the average item mean score.

The SDs for MFI-20 remained roughly parallel for all items, whereas the mean scores varied more across items. SDs ranged from 1.1 to 1.4 and mean scores ranged from 2.1 to 3.9. Four items (#4, 7, 9, and 18) elicited lower response options (indicating less fatigue) by a higher number of respondents. This resulted in a mean score that was >20% lower than the average item mean score (i.e., 3.0). Three items (#8, 16, and 20) had more respondents who preferred the higher response options indicating more fatigue, resulting in a mean score that was >20% higher than the average item mean score.

Targeting

FSS and FIS total scale scores ranged across almost their full possible scoring ranges. The MFI-20 total scores ranged from 27.3 to 93.0 (possible scoring range, 20–100), which suggests that only 82% of the possible scoring range was used. Mean scores for the three fatigue rating scales were fairly close to the scale midpoints (within 1 SD). Floor and ceiling effects were substantially below 20% and skewness was less than ±1 for all three scales (Table 4).

Reliability

Cronbach's α was 0.96 for FSS, 0.99 for FIS, and 0.95 for MFI-20. The results of the test-retest reliability analyses of the three fatigue rating scales are presented in Table 5. All three scales obtained ICC values ≥0.80 and one scale (FIS) yielded an ICC value of 0.90. SEM% ranged from 7% to 10% and SDD% ranged from 20% to 28%, and was the highest (i.e., worst) for FSS. The 95% CI around đ included 0 in all the three fatigue scales.

Convergent validity

There were significant correlations (r_s) between all three fatigue rating scales. The correlation between FSS and FIS was 0.80. The correlations with FSS and FIS total scores for the MFI-20 raw score were 0.79 and 0.80

Table 4. Targeting of the FSS, FIS and MFI-20

	FSS	FIS	MFI-20
Total score	61	51	58
mean±SD (min–max)	4.3±1.8 (1.1–7.0)	54.4±37.8 (0–157)	55.1±13.3 (27.3–93.0)
Skewness	-0.07	0.70	0.69
Floor/ceiling effects (%)	0/6.6	5.9/0	0/0

Total scores are based on participants with complete data at t1. All MFI-20 data are based on transformed interval total scores according to Dencker et al. [19].
FSS, Fatigue Severity Scale; FIS, Fatigue Impact Scale; MFI-20, Multidimensional Fatigue Inventory.
Possible scoring ranges: FSS, 1–7; FIS, 0–160; and MFI-20, 20–100. Greater scores indicate more fatigue.

Table 5. Test-retest reliability of the FSS, FIS and MFI-20

	FSS	FIS	MFI-20
Total score	56	44	52
ICC (95% CI)	0.84 (0.75 to 0.90)	0.90 (0.83 to 0.95)	0.80 (0.67 to 0.88)
đ (95% CI)[a]	0.21 (-0.03 to 0.45)	4.25 (-0.55 to 9.05)	-0.78 (-3.04 to 1.47)
SEM[b] (% of possible scoring range)	0.7 (10)	11.7 (7)	6.0 (7)
SDD[c] (% of possible scoring range)	2.0 (28)	32.3 (20)	16.6 (20)

Total scores are based on participants with complete data at both t1 and t2.
All MFI-20 data are based on transformed interval total scores according to Dencker et al. [19].
FSS, Fatigue Severity Scale; FIS, Fatigue Impact Scale; MFI-20, Multidimensional Fatigue Inventory; SEM, standard error of measurement; SDD, smallest detectable difference.
Possible scoring ranges: FSS, 1–7; FIS, 0–160; and MFI-20, 20–100. Greater scores indicate more fatigue.
[a]đ defined as mean difference in scale scores (*time* 1-*time* 2). [b]Based on ICC, using the formula $SEM=SD_{baseline}\times\sqrt{1-ICC}$.
[c]Based on SEM^2, using the formula $SDD=SEM\times1.96\times\sqrt{2}$.

(p<0.001); the corresponding correlations for the MFI-20 transformed score were 0.47 and 0.49, respectively.

DISCUSSION

Understanding various aspects of the psychometric properties of self-rating scales is a basic, albeit important, starting point when selecting a scale in clinical research. Over the past decade, various strategies were used to evaluate the psychometric properties of self-rating scales. This study is, to the best of our knowledge, the first that includes a comprehensive psychometric evaluation and head-to-head comparison of three fatigue rating scales (FSS, FIS, and MFI-20) in individuals with LEoP.

In summary, our results show that all rating scales displayed acceptable psychometric properties in terms of data completeness, scaling assumptions, targeting and reliability, and high convergent validity. Previous studies have explored various validity and/or reliability aspects of FSS [14-18], FIS [14,15] and MFI-20 [19] in subjects with LEoP by using traditional psychometrics or Rasch analysis. However, the lack of studies evaluating other psychometric aspects limits an adequate comparison of our results with previous studies.

Data completeness was excellent for the three scales, without any missing item responses in FSS and only 0.4%–0.5% missing item responses in MFI-20 and FIS. However, due to the large number of items in FIS (n=40) and the decision not to use imputation for missing responses, 10 subjects (16%) did not report FIS total scores due to one or more missing item response. The developer of FIS states that imputation can be used in cases with less than 10% missing responses [36]. However, the use of imputation is based on assumptions of a participant's response to items, which might be another challenge than the items responded to (which are commonly used as a basis in the imputation) [27], rendering imputation unreliable. Thus, scales with fewer items may be favorable compared with more extensive scales, as this may affect the number of dropouts. In addition, the time needed to respond is another factor determining the selection of a fatigue rating scale.

The items of FSS were roughly parallel in terms of mean scores and SDs, whereas FIS and MFI-20 contained a few items that were rated lower or higher (i.e., indicating less or more fatigue) than the other items. In fact, 12 out of the 40 FIS items were rated as easier and another 10 items were rated as more difficult than the other items. In MFI-20, 4 of the 20 items were rated as easier and another three items were rated as more difficult than the other items. Items within a rating scale are supposed to be "roughly parallel" with the legitimacy of total scores [27,28]. However, no guideline is available describing the limits of parallel items. Item SDs were roughly parallel with items in all three scales and corrected item-total correlations fulfilled the criterion >0.4, which support the use of total scores. Moreover, a previous Rasch analysis of MFI-20 has confirmed its uni-dimensionality and the use of total score [19]. Conversely, a previous Rasch analysis of FSS concluded that a simplified version of the scale (without the first item and with 3 response categories instead of the original 7) is more psychometrically sound than the original scale [18]. Taken together, further studies of these commonly used fatigue scales are required in order to fully establish their construct validity.

All three rating scales appear to be well targeted with very little floor and ceiling effects indicating that the scales can be used to detect changes in fatigue levels in individuals with LEoP. The transformed (Rasch analyzed) score in MFI-20 did not range the full span of possible scale scores (scoring range, 20–100; actual scoring range, 27.3–93.0), which implies that 18% of the scoring range was not used by any participant. However, the corresponding raw scores ranged from 21 to 99 [19], indicating our sample coverage of almost the full possible scoring range.

Reliability coefficients were acceptable for the three scales with Cronbach's α well above the recommended limit of 0.7 [32] consistent with previous studies of Cronbach's α for FSS and FIS [14,16-18]. It is also in agreement with a previous Rasch analysis of MFI-20, which reported the scale's person separation index, considered to be equivalent to Cronbach's α [19]. FIS yielded the highest Cronbach's α (=0.99), suggesting redundant items as Cronbach's α is strongly affected by the length of a rating scale [37]. However, a previous study reported a Cronbach's α of 0.82 for FIS [14], which contradicts this observation.

All three rating scales yielded ICC values >0.70, indicating very good test-retest reliability [21], and can therefore be used to assess fatigue at a group level [32]. Only FIS yielded an ICC of 0.90, which is the lower limit of a rating

scale for individual comparisons [32]. Previous studies have reported an ICC value of 0.91 for FIS [14] and ICC values for FSS ranging from 0.80 to 0.97 [14,16,17]. To the best of our knowledge, no previous study has reported ICC values for MFI-20.

FIS and MFI-20 yielded identical SEM% and SDD%, which implies that a change in scale score in either FIS or MFI-20 greater than 7% of the possible scoring range indicates a real change (above measurement error) on a group level. Correspondingly, a change in the scale score of more than 20% of the possible scoring range indicates a real change (above measurement error) in an individual. The corresponding values for FSS are 10% at the group level (i.e., SEM%) and 28% in an individual (i.e., SDD%).

No systematic differences were detected between the two test occasions in any of the rating scales—95% CI around đ included 0 for all three rating scales—indicating the absence of learning effects.

The correlations between the three scales were high (r_s=0.79–0.80) based on the raw scores, indicating high convergent validity. However, when using the transformed scores for MFI-20 the correlations were lower, most likely as a result of the lower score distribution between the measures. FSS and FIS are aimed at evaluating the impact of fatigue on daily living [11,12]. MFI-20 is intended to assess fatigue 'as experienced by patients' [13]. These constructs appear similar, and our findings suggest that they can be used interchangeably.

Several psychometric properties were similar among the three fatigue rating scales. The differences between them were mainly related to the number of items in the scales. Clearly, a higher number of items yielded increased number of missing responses and signs of item redundancy. Thus, the number of items in a fatigue rating scale is a central factor in determining the choice of scale in clinical investigations of persons with LEoP.

In clinical practice and in previous studies, the total cumulative scores of the three fatigue rating scales were used, even though they were all ordinal scales. Future studies evaluating the construct validity and the unidimensionality of the FIS, using the Rasch method, are needed to determine if the extra 31 items in the FIS are necessary compared with scales carrying fewer items.

Our results are in many ways similar to studies of other neurological conditions. A recent systematic review summarized the psychometric properties (validity and reli-

ability) and clinical utility (ability to detect change) of several fatigue rating scales [38]. The scales were evaluated among people with multiple sclerosis, spinal cord injury, acquired brain injury and Parkinson disease. Overall, the FSS and FIS showed good to excellent reliability (internal consistency and/or test-retest reliability), and acceptable validity and scaling structure with no floor and ceiling effects. The authors suggested that a fatigue measure effective in one condition is not necessarily appropriate for use with another [38]. Therefore, a comprehensive evaluation of the psychometric properties of rating scales for specific conditions is required.

The head-to-head comparison of three commonly used fatigue rating scales using a comprehensive set of analyses is one of the strengths of the study. Furthermore, the high response rate yielded a 'good sample size' for all the analyses, according to the general recommendations [20,21]. The study sample included subjects who were in general moderately bothered by LEoP-related impairments, and the results might vary in persons with a more severe disability. Thus, the inferences of the study should be restricted to patients with moderate LEoP.

The results of this head-to-head comparison suggest that the FSS, FIS, and MFI-20 exhibit sound psychometric properties in terms of data completeness, scaling assumptions, targeting, reliability, and high convergent validity. These results support our hypothesis and indicate that these three scales can be used to assess fatigue in persons with LEoP. However, a scale with fewer items, such as FSS, compared with multiple items may be completed quickly. Further, the risk of missing responses is minimized. Given the similarities and differences between these three scales, the choice of fatigue rating scale in clinical research depends on the research question and the study design.

CONFLICT OF INTEREST

ACKNOWLEDGMENTS

collection. The study was funded by Stiftelsen för bistånd åt rörelsehindrade i Skåne.

REFERENCES

1. Lexell J. Postpoliomyelitis syndrome. In: Frontera WR, Silver JK, Rizzo TD, editors. Essentials of physical medicine and rehabilitation: musculoskeletal disorders, pain, and rehabilitation. Philadelphia: Elsevier Saunders; 2015. p. 775-81.

2. Schanke AK, Stanghelle JK. Fatigue in polio survivors. Spinal Cord 2001;39:243-51.

3. Jensen MP, Alschuler KN, Smith AE, Verrall AM, Goetz MC, Molton IR. Pain and fatigue in persons with post-polio syndrome: independent effects on functioning. Arch Phys Med Rehabil 2011;92:1796-801.

4. Brogardh C, Lexell J. How various self-reported impairments influence walking ability in persons with late effects of polio. NeuroRehabilitation 2015;37:291-8.

5. Lexell J, Brogardh C. Life satisfaction and self-reported impairments in persons with late effects of polio. Ann Phys Rehabil Med 2012;55:577-89.

6. McNalley TE, Yorkston KM, Jensen MP, Truitt AR, Schomer KG, Baylor C, et al. Review of secondary health conditions in postpolio syndrome: prevalence and effects of aging. Am J Phys Med Rehabil 2015;94:139-45.

7. Tersteeg IM, Koopman FS, Stolwijk-Swuste JM, Beelen A, Nollet F; CARPA Study Group. A 5-year longitudinal study of fatigue in patients with late-onset sequelae of poliomyelitis. Arch Phys Med Rehabil 2011;92:899-904.

8. Kalkman JS, Zwarts MJ, Schillings ML, van Engelen BG, Bleijenberg G. Different types of fatigue in patients with facioscapulohumeral dystrophy, myotonic dystrophy and HMSN-I: experienced fatigue and physiological fatigue. Neurol Sci 2008;29 Suppl 2:S238-40.

9. On AY, Oncu J, Atamaz F, Durmaz B. Impact of post-polio-related fatigue on quality of life. J Rehabil Med 2006;38:329-32.

10. Yang EJ, Lee SY, Kim K, Jung SH, Jang SN, Han SJ, et al. Factors associated with reduced quality of life in polio survivors in Korea. PLoS One 2015;10:e0130448.

11. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989;46:1121-3.

12. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. Can J Neurol Sci 1994;21:9-14.

13. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315-25.

14. Oncu J, Atamaz F, Durmaz B, On A. Psychometric properties of fatigue severity and fatigue impact scales in postpolio patients. Int J Rehabil Res 2013;36:339-45.

15. Vasconcelos OM Jr, Prokhorenko OA, Kelley KF, Vo AH, Olsen CH, Dalakas MC, et al. A comparison of fatigue scales in postpoliomyelitis syndrome. Arch Phys Med Rehabil 2006;87:1213-7.

16. Horemans HL, Nollet F, Beelen A, Lankhorst GJ. A comparison of 4 questionnaires to measure fatigue in postpoliomyelitis syndrome. Arch Phys Med Rehabil 2004;85:392-8.

17. Koopman FS, Brehm MA, Heerkens YF, Nollet F, Beelen A. Measuring fatigue in polio survivors: content comparison and reliability of the Fatigue Severity Scale and the Checklist Individual Strength. J Rehabil Med 2014;46:761-7.

18. Burger H, Franchignoni F, Puzic N, Giordano A. Psychometric properties of the Fatigue Severity Scale in polio survivors. Int J Rehabil Res 2010;33:290-7.

19. Dencker A, Sunnerhagen KS, Taft C, Lundgren-Nilsson A. Multidimensional fatigue inventory and postpolio syndrome: a Rasch analysis. Health Qual Life Outcomes 2015;13:20.

20. Hobart JC, Cano SJ, Warner TT, Thompson AJ. What sample sizes for reliability and validity studies in neurology? J Neurol 2012;259:2681-94.

21. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil 2005;84:719-23.

22. Brogardh C, Lexell J, Lundgren-Nilsson A. Construct validity of a new rating scale for self-reported impairments in persons with late effects of polio. PM R 2013;5:176-81.

23. Mattsson M, Moller B, Lundberg IE, Gard G, Bostrom C. Reliability and validity of the Fatigue Severity Scale in Swedish for patients with systemic lupus erythema-

tosus. Scand J Rheumatol 2008;37:269-77.

24. Flensner G, Lindencrona C. The cooling-suit: case studies of its influence on fatigue among eight individuals with multiple sclerosis. J Adv Nurs 2002;37:541-50.

25. Smets EM, Garssen B, Cull A, de Haes JC. Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. Br J Cancer 1996;73:241-5.

26. Furst CJ, Ahsberg E. Dimensions of fatigue during radiotherapy: an application of the Multidimensional Fatigue Inventory. Support Care Cancer 2001;9:355-60.

27. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009;13:1-177.

28. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 1998;51:945-52.

29. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. Health Technol Assess 2004;8:1-48.

30. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.

31. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res 2004;13:571-86.

32. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193-205.

33. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 5th ed. Oxford: Oxford University Press; 2015.

34. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.

35. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. 5th ed. Boston: Houghton Mifflin; 2003.

36. Fisk JD. Scaling and scoring of the Fatigue Impact Scale version 2.0 (FIS). Lyon, France: MAPI Research Trust; 2009.

37. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess 2003;80:99-103.

38. Tyson SF, Brown P. How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far. Clin Rehabil 2014;28:804-16.