

A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model

Annors Lerdal · Anders Kottorp · Caryl Gay ·
Bradley E. Aouizerat · Carmen J. Portillo ·
Kathryn A. Lee

Accepted: 11 February 2011 / Published online: 10 March 2011
© Springer Science+Business Media B.V. 2011

Abstract

Purpose To examine the psychometric properties of the 9-item Fatigue Severity Scale (FSS) using a Rasch model application.

Methods A convenience sample of HIV-infected adults was recruited, and a subset of the sample was assessed at 6-month intervals for 2 years. Socio-demographic, clinical, and symptom data were collected by self-report questionnaires. CD4 T-cell count and viral load measures were obtained from medical records. The Rasch analysis included 316 participants with 698 valid questionnaires.

Results FSS item 2 did not advanced monotonically, and items 1 and 2 did not show acceptable goodness-of-fit to the Rasch model. A reduced FSS 7-item version demonstrated acceptable goodness-of-fit and explained 61.2% of

the total variance in the scale. In the FSS-7 item version, no uniform Differential Item Functioning was found in relation to time of evaluation or to any of the socio-demographic or clinical variables.

Conclusion This study demonstrated that the FSS-7 has better psychometric properties than the FSS-9 in this HIV sample and that responses to the different items are comparable over time and unrelated to socio-demographic and clinical variables.

Keywords Fatigue · HIV · Psychometrics · Symptoms · Quality of life

Abbreviations

AIDS	Acquired immune deficiency syndrome
BMI	Body mass index
DIF	Differential item functioning
DS	Daytime sleepiness subscale
FSS	Fatigue severity scale
GSDS	General sleep disturbance scale
HIV	Human immunodeficiency virus
PCA	Principal component analysis
SE	Standard error

A. Lerdal (✉)

Lovisenberg Deaconal University College,
Lovisenbergveien 15b, 0456 Oslo, Norway
e-mail: Annors.Lerdal@ldh.no

A. Kottorp

Division of Occupational Therapy, Department of Neurobiology,
Care Sciences and Society, Karolinska Institutet, Stockholm,
Sweden

C. Gay · K. A. Lee

Department of Family Health Care Nursing, University
of California, San Francisco, San Francisco, CA, USA

B. E. Aouizerat

Department of Physiological Nursing and Institute for Human
Genetics, University of California, San Francisco,
San Francisco, CA, USA

C. J. Portillo

Community Health Systems, University of California,
San Francisco, San Francisco, CA, USA

Introduction

Fatigue is a distressing and frequent symptom among people with HIV and is related to lower quality of life [1, 2]. Individuals describe fatigue as silent and invisible, which can make it difficult for health care providers to understand how seriously it impacts patients' lives [3]. Prevalence estimates for fatigue range from 37 to 65% [4–7]. Several studies of people with HIV have found that fatigue is related to

psychosocial factors such as sleep disturbance [4, 8], depressive symptoms [9, 10], anxiety [9], or post-traumatic stress [11]. Studies of physiological correlates of fatigue report conflicting findings about relationships between fatigue and CD4+ T-cell count [12, 13] or viral load [6, 7, 11, 14, 15].

Since fatigue is a subjective phenomenon, researchers and clinicians rely on subjective measures to indicate need for intervention or effectiveness of treatment. It is therefore important to have reliable and valid instruments to measure fatigue as a symptom that can vary over time or across illness severity.

The fatigue severity scale

The fatigue severity scale (FSS) [16] is one of the most frequently used fatigue measures for adults with chronic illness, including multiple sclerosis [17] and stroke [18]. It has also been used in studies of people with HIV [10, 19–21] and in the general population [22]. In a recent published review of 18 fatigue instruments used in chronic illness research, the FSS was rated highest on robust psychometric properties [23]. An important advantage of this instrument is that it is short and consists of only 9 items. This can be especially important for people with limited stamina in completing lengthy questionnaires. Furthermore, normative data from the general population are available. Clinical cut-off scores have also been reported, but to our knowledge, the different cut-off values have not been clinically validated. The nine items are formulated as statements (see Table 1), with seven items related to fatigue interference, one item related to the experience of fatigue itself (item 3), and one item about what causes fatigue (item 2).

The aim of this study was to examine the psychometric properties of the FSS in a sample of HIV-infected adults and to propose a revised version with better psychometric properties. The specific objectives were to determine: (1) the fit of the items to the Rasch model and unidimensionality, (2)

person-response validity, (3) reliability, (4) concurrent validity, and (5) the presence of uniform differential item functioning (DIF) in relation to time intervals for evaluation, socio-demographic factors, and clinical variables.

Methods

Data were collected as part of a prospective longitudinal study of adults with HIV [19]. The study was designed to characterize the symptom experience of HIV-positive adults and identify biological and genetic markers of their symptom experience.

Sample and procedures

A convenience sample of 350 adults with HIV was enrolled in the study over a 3-year period (April 2005 to December 2007). The participants were recruited using flyers posted at local HIV clinics and community sites. Study visits were conducted at the University of California, San Francisco, Clinical Research Center. Eligible participants were English-speaking, at least 18 years old, and diagnosed with HIV at least 30 days before enrollment. Individuals were excluded if they currently used illicit drugs, worked nights, had been pregnant in the previous 3 months, or reported having a diagnosed sleep disorder, bipolar disorder, schizophrenia, or dementia.

Data collection and measurements

All participants completed a baseline assessment. Those not already reporting significant sleep disturbance or fatigue at baseline were included in a longitudinal study focused on the development of these symptoms over time and were assessed at 6-month intervals for up to 2 years, for a maximum of five assessments each. Self-report questionnaires were used to collect data on demographics,

Table 1 The fatigue severity scale (FSS); English (US) version

A. Items

1. My motivation is lower when I am fatigued
2. Exercise brings on my fatigue
3. I am easily fatigued
4. Fatigue interferes with my physical functioning
5. Fatigue causes frequent problems for me
6. My fatigue prevents sustained physical functioning
7. Fatigue interferes with carrying out certain duties and responsibilities
8. Fatigue is among my three most disabling symptoms
9. Fatigue interferes with my work, family, or social life

B. Response categories on a Likert scale

Strongly disagree							Strongly agree	
1	2	3	4	5	6	7		

clinical characteristics, and concurrent symptoms. CD4+ T-cell count and viral load measures were obtained from the most recent laboratory report in their medical record.

Variables used for analysis

The questionnaire included demographic information on age, gender, race/ethnicity, level of education, partner status, and employment status. Participants also reported whether they had ever received a diagnosis of AIDS. Measures of height and weight were obtained from each participant and used to calculate body mass index (BMI).

Fatigue during the last week was measured with the FSS, a 9-item unidimensional questionnaire developed by Krupp et al. [16] (Table 1). Each item is scored on a 7-point Likert scale ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). The mean score of the 9 items is used to estimate fatigue severity.

The *Daytime Sleepiness (DS) subscale* of the General Sleep Disturbance Scale (GSDS) [24] was used to assess impact of sleepiness on daytime function. The subscale consists of 7 symptoms (e.g., struggle to stay awake during the day). Respondents rate how often they experienced each item in the past week using a numeric rating scale ranging from 0 (“no days”) to 7 (“every day”). The item scores are averaged to yield a subscale score ranging from 0 to 7. A score of 3 or higher is used to identify those experiencing impaired daytime function due to sleepiness at least 3 days per week. The Cronbach’s alpha coefficient of the subscale was 0.77 in this sample.

A 4-item numerical rating version of the *Visual Analog Scale for Fatigue* [25] was used to prospectively measure fatigue severity each morning and evening for three consecutive days. Because morning fatigue ratings are likely to be confounded by sleep disturbance, only evening fatigue was used for this analysis. Each item was rated on a scale from 0 to 10 and averaged to obtain a mean score for each person. The Cronbach alpha coefficient of the 4-item evening fatigue scale was 0.91 in this sample.

Statistical analysis

A Rasch model was chosen to analyze the FSS for two reasons. First, items in the FSS represent different aspects of fatigue in everyday life that are assumed to vary in severity among adults with HIV infection. The Rasch model takes each item scored and adjusts the final person measure based on relative differences in item severity. Secondly, Rasch models are suitable for handling data where items may be missing at random. Even though only 16 of the 2,844 (0.6%) item scores were missing among the 316 participants, we did not have to exclude any participant due to missing values by the use of a Rasch model [26–28].

The WINSTEPS analysis software program, version 3.69.1.16 [29], was used to conduct the Rasch analysis. A Rasch analysis first converts raw item scores from a questionnaire into equal-interval measures using a logarithmic transformation of the odds probabilities of responses. These converted values are then used to examine whether items from a scale measure a unidimensional construct, viewed as crucial in both classical and modern test statistics [26, 30]. The transformation simultaneously results in an estimation of a person’s fatigue measure as well as difficulty of the items along a calibrated continuum (from easy to harder to agree with). Rasch models are probabilistic and based on theoretical assertions against which the actual pattern of responses is validated. Although the FSS uses a generic rating scale from 1 to 7, it may not function in a similar manner across all items. Therefore, a partial credit model, developed for scales where ratings may differ across items, was applied to the FSS in this sample.

The psychometric properties of the rating scale used in the FSS were initially evaluated (step 1) with the following criteria: (a) average measures for each step category on each item should advance monotonically, and (b) a criterion less than 2.0 was expected in outfit mean square (*MnSq*) values for step category calibrations [31, 32]. The fit of the items to the Rasch model was then analyzed (step 2) followed by principal component analysis to address unidimensionality (step 3), aspects of person-response validity (step 4) and person-separation reliability (step 5). Finally, differential item functioning (DIF) analyses were performed to further support the fit of the items to the Rasch model (step 6).

Evidence of *internal-scale validity* (step 2) and *person-response validity* (step 4) were investigated using item and person goodness-of-fit statistics using the WINSTEPS program to generate mean square (*MnSq*) residuals and standardized *z*-values. These indicate the degree of match between actual responses on the FSS and expected responses from the Rasch model. The goodness-of-fit statistics were evaluated using both *infit* and *outfit* statistics. *Infit statistics* are information-weighted fit statistics that give relatively more weight to the performances of persons who are well targeted to the item difficulty calibrations. *Outfit statistics* are not weighted and therefore are more sensitive to outlying scores. As *infit* statistics are more informative when exploring the fit of the items to the Rasch model and person-response validity [33, 34], we chose *infit* statistics to evaluate goodness-of-fit across individual items and across persons in this study.

The *MnSq* fit statistic has an expected value of 1.0 and is preferable for item goodness-of-fit with polytomous data (as in the FSS), as it is less sensitive to sample size compared to *z* [35]. We therefore chose to use a sample size-adjusted criterion [35] for *item goodness-of-fit* set for *infit MnSq* values between 0.7 and 1.3 logits.

The final criterion for evaluating *person goodness-of-fit* was to accept infit *MnSq* values ≤ 1.4 logit with an associated z value < 2 [36, 37]. It is generally accepted that 5% of the sample, by chance, may not demonstrate acceptable goodness-of-fit without a serious threat to person-response validity [36, 37].

To minimize risk of additional explanatory factors, a *principal component analysis* (PCA) of residuals was performed to evaluate unidimensionality of the FSS (step 3) [38]. Two criteria were set: (1) at least 50% of the total variance should be explained by the first latent variable (perceived fatigue), and (2) any additional factor should explain $< 5\%$ of the remaining variance of residuals (with an associated eigenvalue ≤ 1.4) after removal of the first latent variable [39, 40].

To further determine whether the FSS could distinguish people with different levels of fatigue, *person-separation reliability* was investigated with the individual and group *standard error (SE)* of measures (step 5). For a scale to distinguish between three or more groups, a *person-separation index* of 2.0 is required. For the purpose of comparison to more traditional reliability estimates, the Rasch-equivalent Cronbach alpha statistic was also reported.

Initially, an analysis of all 9 FSS items was performed. If an item did not demonstrate acceptable goodness-of-fit to the model according to criteria, one item at a time was removed and psychometric properties were reanalyzed with the remaining items. This procedure was repeated until all items demonstrated acceptable goodness-of-fit. After each item removal, unidimensionality, person-response validity, and reliability of the FSS measures were re-evaluated as described above.

A number of differential item functioning (*DIF*) analyses were also performed to evaluate stability of response patterns in relation to repeated time intervals, socio-demographic factors, and clinical variables (step 6). The magnitude of DIF was evaluated using the Mantel–Haenszel statistic for polytomous scales using log-odds estimators [41, 42] in the WINSTEPS program (1% alpha with Bonferroni correction). Given the impact of sample size on standard errors of item difficulty estimates, differences between items could be artificially significant but not clinically relevant; therefore, we also evaluated the size of the discrepancy between item difficulty estimates. This was evaluated using an additional approach in which the item standard error was set at 0.15 logit, indicating that an item difference must exceed 0.43 logit in order to be clinically relevant [43–45]. For the FSS items to be considered stable across external variables, no item should have a significant or clinically relevant DIF.

SPSS for Windows Version 14.0 software (SPSS Inc., IL, USA) was used to analyze demographic data, concurrent validity, and potential differences in demographic or clinical variables between persons with and without misfit.

Categorical data were analyzed using chi-square. *Concurrent validity* was assessed by correlations (Spearman's rho) between the Rasch generated FSS measures, daytime sleepiness scores, and evening fatigue severity ratings. Cronbach alpha coefficients were reported for the daytime sleepiness scale and the evening fatigue severity scale.

Ethics

The study was approved by the Committee on Human Research at the University of California, San Francisco. All participants provided written informed consent.

Results

Sample characteristics

After excluding 34 participants for testing positive for an illicit drug or having no FSS data, the final sample included 316 participants. Mean age at baseline was 45 years ($SD \pm 8.4$). The sample characteristics are described in Table 1. A total of 698 FSS questionnaires were included in the analysis (316 at baseline, 116 at 6 months, 103 at 12 months, 90 at 18 months, and 73 at 24 months). Initial analyses were conducted on the baseline sample ($N = 316$). To determine whether there was differential item functioning in relation to multiple time points of evaluation, the analyses were repeated with the complete longitudinal ($N = 698$) as part of step 6 (Table 2).

Rating scale functioning (step 1)

When evaluating rating scale function of the FSS-9 for the baseline sample, items 1 and 2 did not meet the set criteria. The average step calibration measures did not advance monotonically for item 2, and both items were associated with higher than acceptable outfit *MnSq* values (Table 3). The other 7 items demonstrated acceptable values. Analysis proceeded keeping items 1 and 2 in the FSS to evaluate further aspects of FSS-9 validity for the HIV population.

The fit of the items to the Rasch model and unidimensionality (steps 2 and 3)

In the analysis of the 9-item FSS, items 1 and 2 did not demonstrate acceptable goodness-of-fit. Analysis continued by removing the item demonstrating the least acceptable goodness-of-fit (i.e., item 2) and repeating the analysis on the remaining items. The item removal process continued until all remaining items had acceptable goodness-of-fit. Subsequent iterations also removed item 1 (see Table 3), and the seven remaining items all demonstrated acceptable

Table 2 Demographic and clinical characteristics of the baseline sample ($N = 316$)

Variables	Total sample N (%)
<i>Demographic variables</i>	
Age groups	
22–39 years	86 (27.2)
40–49 years	133 (42.1)
50–77 years	97 (30.7)
Gender	
Men	216 (68.4)
Women	77 (24.4)
Transgender	23 (7.3)
Race/ethnicity	
Black/African American	122 (38.6)
White/Caucasian	129 (40.8)
Other	65 (20.6)
Level of formal education	
Some high school or less	47 (14.9)
High school diploma or equivalent	94 (29.7)
Some college	107 (33.9)
College degree or higher	68 (21.5)
Partner status: In a relationship	109 (34.5)
Employment	
Not working and not in school	267 (84.5)
Employed or in school	49 (15.5)
<i>Clinical variables</i>	
AIDS diagnosis: yes	165 (52.2)
Body mass index ($n = 313$)	
<24.9	131 (41.9)
≥ 25.0	182 (58.1)
CD4+ T-cell count ($n = 312$)	
<200 cells/mm ³	54 (17.3)
Viral load ($n = 294$)	
≥ 10 000 copies/ml	58 (19.7)

goodness-of-fit. The continuum of difficulty calibrations for the remaining 7 FSS items is presented in Table 4.

The PCA for the final 7-item FSS version suggests that the Rasch model aligns with measuring fatigue interference. The Rasch model explained 61.2% of the total variance in the dataset, which was above the 50% criterion and further supports unidimensionality, in addition to item fit. The secondary factor explained an additional 8.5% of the variance, which was slightly higher than the expected 5%. Therefore, evidence of unidimensionality is still mixed in the 7-item FSS version.

Person-response validity and reliability (steps 4 and 5)

Of the 316 FSS surveys, 291 (92.1%) demonstrated acceptable goodness-of-fit to the Rasch model. Person misfit occurred in 25 persons, and therefore FSS-7 demonstrated a somewhat higher level of misfit among participants than expected. A comparison of responses with misfit and without misfit showed a higher proportion of persons with misfit scores among black/African Americans ($n = 19$, 15.6%) than among Caucasians ($n = 5$, 3.9%) or other ethnic groups ($n = 1$, 1.5%, $\chi = 16.3$, $P < 0.001$).

The number of participants with maximum and minimum scores (ceiling and floor effects) across the different FSS item solutions is shown in Table 3. The proportion of participants demonstrating maximum or minimum scores was higher for the FSS-7 compared to the FSS-9 version.

The person-separation index in the FSS-7 was 2.341, indicating that the Rasch model captures fatigue from the dimension of its impact or interference and can detect more than three statistically distinct groups of participants within the sample. The Rasch-equivalent Cronbach alpha coefficient for the FSS-9 was 0.90, and the alpha for the FSS-7 was 0.92.

In Fig. 1, the distributions of persons' ability measures and item calibration values (including each item threshold

Table 3 Presentation of the outcomes from the first five steps in Rasch analysis of the psychometric properties of the different fatigue severity scale (FSS) item solutions in adults with HIV infection

Step		FSS 9-items ($N = 316$)	FSS 8-items Item 2 removed ($N = 316$)	FSS 7-items Items 1 and 2 removed ($N = 316$)	FSS 7-items Items 1 and 2 removed ($N = 698$)
1	Items not meeting criteria for rating scale	1,2	1	None	None
2	Item misfit	1,2	1	None	3
3	Variance explained, %	55.2%	58.9%	61.2%	58.8%
	2nd dimension, %	9.0%	8.1%	8.5%	9.1%
4	Person misfit, n (%)	29 (9.2)	25 (7.9)	25 (7.9)	58 (8.3)
	Maximum score, n (%)	5 (1.6)	10 (3.2)	10 (3.2)	14 (2.0)
	Minimum score, n (%)	18 (5.7)	19 (6.0)	26 (8.2)	75 (10.7)
5	Person-separation index (without extremes)	2.09	2.24	2.31	2.08

Table 4 FSS-7 item hierarchy for adults with HIV infection

	Measure (logit)	SE (logit)	Item
Harder to agree with	0.30	0.05	Item 5: Fatigue causes frequent problems for me
	0.17	0.05	Item 6: My fatigue prevents sustained physical functioning
	0.13	0.05	Item 3: I am easily fatigued
	0.03	0.05	Item 7: Fatigue interferes with carrying out certain duties and responsibilities
	-0.02	0.05	Item 9: Fatigue interferes with my work, family, or social life
	-0.16	0.05	Item 8: Fatigue is among my three most disabling symptoms
Easier to agree with	-0.47	0.05	Item 4: Fatigue interferes with my physical functioning

per category score) are presented. Most of the fatigue interference among clients with HIV was captured within the FSS-7 version, despite earlier mentioned ceiling and floor effects (see also Table 3).

Presence of uniform DIF in the FSS-7 in relation to demographic factors, clinical variables, and multiple time points of evaluation (step 6)

There was no uniform DIF among the 7 items by age groups in the baseline sample (22–39, 40–49, 50–77 years), gender (men, women, transgender), employment (yes, no), or partner status (in a relationship or not). Item 7 demonstrated significant uniform DIF in relation to race/ethnicity (black/African American, white/Caucasian, other), but it was not a clinically relevant difference. Items 5, 7, and 8 also demonstrated significant uniform DIF for level of education (some high school or less, high school diploma or equivalent, some college, college degree or higher), but again differences were not clinically relevant. There was also no uniform DIF among any clinical variable (i.e. AIDS diagnosis [yes, no], BMI [$<24.9 \geq 25.0$], CD4+ T-cell count [<200 cells/mm³, ≥ 200 cells/mm³], viral load [$<10,000$ copies/ml, $\geq 10,000$ copies/ml]). With no evidence of relevant clinical differences, the uniform DIF in the FSS-7 did not require further action (e.g., item split techniques).

To ensure that the findings in relation to aspects of validity in the FSS-7 were stable even when including multiple measures per person, we repeated all Rasch analyses of the FSS-7 using data from all available time-points ($n = 698$) (see Table 3). Because no notable differences were observed between analyses with the 698 records and the 316 independent baseline measures, it was concluded that no items in the FSS-7 demonstrated uniform DIF in relation to multiple time points of evaluation.

Concurrent validity

Adequate concurrent validity was demonstrated with bivariate relationships at baseline ($N = 316$) between the

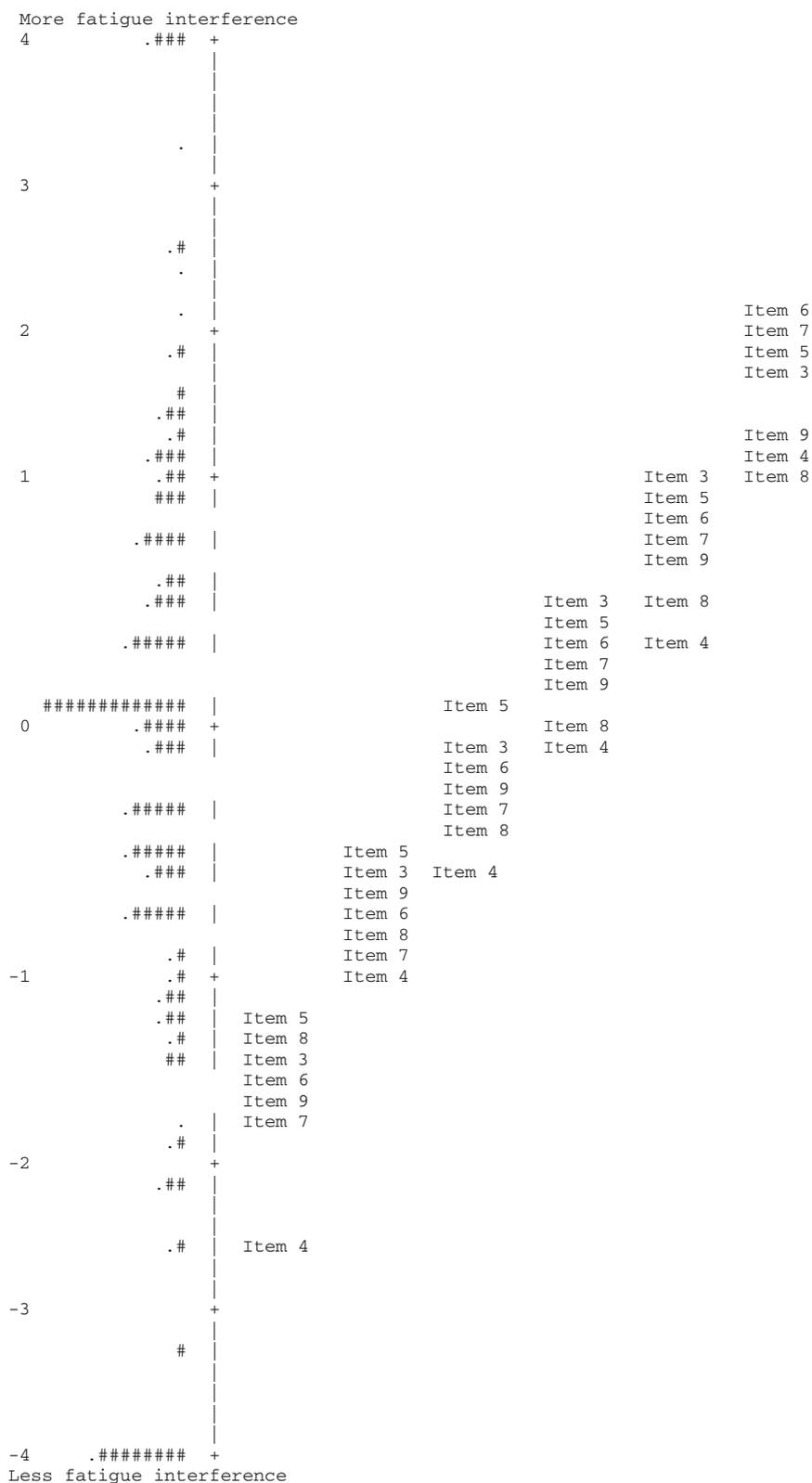
FSS the Daytime Sleepiness subscale from the GSDS; Spearman's $\rho = 0.53$ for the FSS-7 and $\rho = 0.52$ for the FSS-9. The Spearman's correlation coefficient between FSS and evening fatigue was 0.37 for the FSS-7 and 0.38 for the FSS-9.

Discussion

Results of this Rasch analysis suggest that a 7-item FSS has better psychometric properties, overall, than the original 9-item FSS for adults living with HIV. The FSS-7 has satisfactory fit of the items to the Rasch model, stable response patterns over time, no uniform DIF in relation to any clinical variables, and no clinically relevant DIF in relation to any of the socio-demographic variables. The items in the FSS-7 are all statements related to fatigue's interference with daily functioning, since items related to the cause for fatigue and experience of fatigue were excluded in the FSS-7. Thus, we conclude that the FSS-7 measures fatigue interference or impact as a unidimensional concept. Several other studies have tested the unidimensionality of the FSS-9 and concluded that it is a unidimensional scale [23, 46]. Since fatigue is theoretically understood as a multi-dimensional concept [47, 48], including dimensions such as distress, and mental and physical fatigue [47, 49], future studies of fatigue associated with HIV infection should use different measures in order to capture its different dimensions (Whitehead 2009). Like the FSS-9, the explained variance of the latent variable in the FSS-7 was satisfactory, although for both the FSS-7 and the FSS-9, the proportion of variance explained by the second dimension was above the expected 5%.

Results from this study are fairly consistent with those of two recent psychometric studies evaluating the FSS among people with multiple sclerosis using Rasch analysis [50, 51]. Due to inconsistent responding patterns, both studies concluded that exclusion of items 1 and 2 would improve the fit of the items to the Rasch model. In our

Fig. 1 Item/Person map with FSS-7 with each item category threshold calibration value (1–2, 2–3, 3–4, 4–5, 5–6, and 6–7)



sample of adults with HIV, no additional items had to be excluded in order to meet criteria of item fit. [51]. Based on findings from these studies, we conclude that in order to

measure fatigue’s impact or interference with a valid measure across different diagnostic groups, items 1 and 2 should always be excluded from calculation of the mean

score, but additional items to be excluded may vary by type of clinical population.

The item hierarchy for the FSS-7 demonstrated a stable pattern over time for the HIV sample and provides support for the validity of comparisons over time. When analysis of all time points was repeated and findings were replicated with an analysis of baseline data only, use of repeated measures did not impact our final conclusions regarding validity of the FSS-7.

When evaluating the item hierarchy of the FSS, item 4 (“fatigue interferes with my physical functioning”) and item 8 (“fatigue is among my three most disabling symptoms”) were the items most easily agreed with, while item 5 (“fatigue causes frequent problems for me”) was the most difficult for subjects to agree with. Among the seven items in the FSS-7, item 8 is the most general and abstract statement. Thus, adults with HIV in our sample found it relatively easy to report fatigue as one of their most disabling symptoms. This may reflect the significant impact fatigue has on quality of life, even considering the many other symptoms they experience [52]. To understand whether this finding is unique to the HIV population, additional research is needed to determine whether similar item hierarchies and distributions along the continuum of fatigue are observed in other clinical populations.

The mixed findings regarding unidimensionality of both the FSS-9 and FSS-7 suggest that additional consideration be given to what the scale actually measures. Fatigue is still a poorly defined construct for persons with HIV/AIDS, and some FSS items address impact of fatigue or how it interferes with daily activity rather than fatigue severity. While these concepts are likely strongly related, it is possible to imagine instances in which they might differ. For example, people with demanding lives and multiple roles may find that the same fatigue severity has a greater impact on their lifestyle than someone with fewer social, occupational, or other demands. Further studies are needed to explore and define the construct of fatigue interference for this sample, using a variety of both qualitative and quantitative approaches.

Our study showed that black/African Americans in the sample were more likely to have misfit, indicating that this group might experience fatigue in a systematically different way from other ethnic groups. Previous studies [52, 53] suggest that black/African Americans may report less symptom burden than expected, possibly due to genetic, socio-cultural, or spiritual factors. A recent published study [51] reported DIF of some of the FSS items between a Norwegian and a Swedish cohort. Future studies should evaluate potential DIF in relation to race/ethnicity, possibly using item split techniques to make valid comparisons between groups and assess whether these differences have any clinical relevance. Why race/ethnicity would result in

differences with the FSS or how race/ethnicity would interact with other socio-demographic and clinical variables would require larger groups for more in-depth analyses.

Both FSS-9 and FSS-7 had high estimates of internal consistency assessed with Cronbach’s alpha. However, our Rasch analysis was able to detect problems with unidimensionality that the Cronbach alpha coefficients could not. Future studies should therefore be more cautious in the use and interpretation of Cronbach alpha coefficients in relation to aspects of validity of clinical assessments. The bivariate analysis assessing concurrent validity of the FSS-7 showed low to moderate correlations and indicated that the FSS-7 measure of fatigue interference is related to, but not the same as Daytime Sleepiness or evening fatigue severity. Different definitions, assessments, and approaches should therefore be used in order to target the clinically relevant aspect of fatigue addressed. The relationships between such different aspects must however in the future be both theoretically clarified and empirically tested. The outcomes of this study indicate that the FSS-7 could be used in a valid manner to measure and compare groups of adults living with HIV with variations in *how fatigue impacts or interferes with functioning* over time and across socio-demographics and clinical variables without any threat to validity.

This study had a number of limitations. First, the sample used in this study was a convenience sample and may not be representative of the larger population of adults living with HIV. Second, the analysis of differential item functioning in relation to multiple time points was limited by the subsequent exclusion of participants reporting sleep disturbance or fatigue at their baseline measures. This limitation was a characteristic of the original study, but future studies should include multiple measurements from participants experiencing the full range of fatigue symptoms.

Conclusion

The 7-item version of the FSS derived from the Rasch analysis had better psychometric properties than the FSS-9 in this sample of adults living with HIV infection. With fewer items, it remains a valid and reliable measure of fatigue, while losing very little of the descriptive power of the FSS-9. However, additional research is needed to determine whether these findings are consistent across clinical population, across culture or race/ethnicity, and across other relevant demographic and clinical variables. Since the FSS has been rated one of the best fatigue measures in terms of its psychometric properties, results from this study suggest that other symptoms and quality of life measures might benefit from further examination using a Rasch model as well.

Acknowledgments This research was supported by a grant from the National Institute of Mental Health (NIMH, 5 R01 MH074358). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health. Data collection was supported by the General Clinical Research Center in the UCSF CTSA (1 UL RR024131). Dr. Lerdal has received funding from the Research Council of Norway (Grant # 19256), the Norwegian Nurses Organization and the U.S.–Norway Fulbright Foundation. Dr. Aouizerat is supported by an NIH Roadmap K12 (KL2 RR024130). Authors wish to acknowledge the contributions to the study from Traci Coggins, Skip Davis, Ryan Kelly, Yeonsu Song, Kristen Nelson, and Matthew Shullick.

References

- Cunningham, W. E., Shapiro, M. F., Hays, R. D., Dixon, W. J., Visscher, B. R., George, W. L., et al. (1998). Constitutional symptoms and health-related quality of life in patients with symptomatic HIV disease. *American Journal of Medicine*, *104*, 129–136.
- Marcellin, F., Preau, M., Ravaux, I., Dellamonica, P., Spire, B., & Carrieri, M. P. (2007). Self-reported fatigue and depressive symptoms as main indicators of the quality of life (QOL) of patients living with HIV and Hepatitis C: Implications for clinical management and future research. *HIV Clinical Trials*, *8*, 320–327.
- Jenkin, P., Koch, T., & Kralik, D. (2006). The experience of fatigue for adults living with HIV. *Journal of Clinical Nursing*, *15*, 1123–1131.
- Lee, K. A., Portillo, C. J., & Miramontes, H. (2001). The influence of sleep and activity patterns on fatigue in women with HIV/AIDS. *Journal of the Association of Nurses in Aids Care*, *12*(Suppl), 19–27.
- Phillips, K. D., Sowell, R. L., Rojas, M., Tavakoli, A., Fulk, L. J., & Hand, G. A. (2004). Physiological and psychological correlates of fatigue in HIV disease. *Biological Research for Nursing*, *6*, 59–74.
- Henderson, M., Safa, F., Easterbrook, P., & Hotopf, M. (2005). Fatigue among HIV-infected patients in the era of highly active antiretroviral therapy. *HIV Medicine*, *6*, 347–352.
- Sullivan, P. S., & Dworkin, M. S. (2003). Prevalence and correlates of fatigue among persons with HIV infection. *Journal of Pain and Symptom Management*, *25*, 329–333.
- Pence, B. W., Barroso, J., Leserman, J., Harmon, J. L., & Salahuddin, N. (2008). Measuring fatigue in people living with HIV/AIDS: Psychometric characteristics of the HIV-related fatigue scale. *AIDS Care*, *20*, 829–837.
- Barroso, J., Preisser, J. S., Leserman, J., Gaynes, B. N., Golden, R. N., & Evans, D. N. (2002). Predicting fatigue and depression in HIV-positive gay men. *Psychosomatics*, *43*, 317–325.
- Millikin, C. P., Rourke, S. B., Halman, M. H., & Power, C. (2003). Fatigue in HIV/AIDS is associated with depression and subjective neurocognitive complaints but not neuropsychological functioning. *Journal of Clinical and Experimental Neuropsychology*, *25*, 201–215.
- Barroso, J., Hammill, B. G., Leserman, J., Salahuddin, N., Harmon, J. L., & Pence, B. W. (2010). Physiological and psychosocial factors that predict HIV-related fatigue. *AIDS Behavior*.
- Darko, D. F., McCutchan, J. A., Kripke, D. F., Gillin, J. C., & Golshan, S. (1992). Fatigue, sleep disturbance, disability, and indices of progression of HIV infection. *American Journal of Psychiatry*, *149*, 514–520.
- Walker, K., McGown, A., Jantos, M., & Anson, J. (1997). Fatigue, depression, and quality of life in HIV-positive men. *Journal of Psychosocial Nursing and Mental Health Services*, *35*, 32–40.
- Simmonds, M. J., Novy, D., & Sandoval, R. (2005). The differential influence of pain and fatigue on physical performance and health status in ambulatory patients with human immunodeficiency virus. *Clinical Journal of Pain*, *21*, 200–206.
- Voss, J. G. (2005). Predictors and correlates of fatigue in HIV/AIDS. *Journal of Pain and Symptom Management*, *29*, 173–184.
- Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*, *46*, 1121–1123.
- Johansson, S., Ytterberg, C., Hillert, J., Widen, H. L., & von Koch, L. (2008). A longitudinal study of variations in and predictors of fatigue in multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, *79*, 454–457.
- Lerdal, A., Bakken, L. N., Kouwenhoven, S. E., Pedersen, G., Kirkevold, M., Finset, A., et al. (2009). Poststroke fatigue—a review. *Journal of Pain and Symptom Management*, *38*, 928–949.
- Aouizerat, B. E., Miaskowski, C. A., Gay, C., Portillo, C. J., Coggins, T., Davis, H., et al. (2010). Risk factors and symptoms associated with pain in HIV-infected adults. *Journal of the Association of Nurses AIDS Care*, *21*, 125–133.
- Rabkin, J. G., McElhiney, M. C., Rabkin, R., & Ferrando, S. J. (2004). Modafinil treatment for fatigue in HIV+ patients: A pilot study. *Journal of Clinical Psychiatry*, *65*, 1688–1695.
- Rabkin, J. G., McElhiney, M. C., Rabkin, R., & McGrath, P. J. (2010). Modafinil treatment for fatigue in HIV/AIDS: A randomized placebo-controlled study. *Journal of Clinical Psychiatry*.
- Lerdal, A., Wahl, A., Rustoen, T., Hanestad, B. R., & Moum, T. (2005). Fatigue in the general population: A translation and test of the psychometric properties of the Norwegian version of the fatigue severity scale. *Scandinavian Journal of Public Health*, *33*, 123–130.
- Whitehead, L. (2009). The measurement of fatigue in chronic illness: A systematic review of unidimensional and multidimensional fatigue measures. *Journal of Pain and Symptom Management*, *37*, 107–128.
- Lee, K. A. (1992). Self-reported sleep disturbances in employed women. *Sleep*, *15*, 493–498.
- Lee, K. A., Hicks, G., & Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Research*, *36*, 291–298.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model. Fundamental measurement in the human sciences*. Mahawah, New Jersey: Erlbaum Publishers.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Linacre, J. M. (2006). *Winsteps computer program, Version 3.63.0*. Chicago: Rasch Measurement.
- Linacre, J. M. (2010). *Winstep—Rasch model computer program*. Version 3.69.1.16. <http://www.winsteps.com>, 2010 Retrieved June 2010.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Beverly Hills and London: Sage Publications, Inc.
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models and applications* (pp. 258–278). Maple Grove: JAM Press Publisher.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*, 85–106.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

34. Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.
35. Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8, 33.
36. Nilsson, I., & Fisher, A. G. (2006). Evaluating leisure activities in the oldest old. *Scandinavian Journal of Occupational Therapy*, 13, 31–37.
37. Patomella, A. H., Tham, K., & Kottorp, A. (2006). P-drive: assessment of driving performance after stroke. *Journal of Rehabilitation Medicine*, 38, 273–279.
38. Linacre, J. M. (2005). *Rasch-model computer programs*. Internet 2005. Retrieved June 5, 2007, available from: URL: <http://www.winsteps.com>.
39. Smith, R. M., & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice*. Greenwich: Ablex.
40. Raiche, G. (2005). Critical eigenvalue sizes in standardized residual principal component analysis. *Rasch Measurement Transactions*, 19, 1012.
41. Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel Haenszel procedure. *Journal of American Statistical Association*, 58, 690–700.
42. Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
43. Petersson, I., Lilja, M., Hammel, J., & Kottorp, A. (2008). Impact of home modification services on ability in everyday life for people ageing with disabilities. *Journal of Rehabilitation Medicine*, 40, 253–260.
44. Kottorp, A., Bernspang, B., & Fisher, A. G. (2003). Activities of daily living in persons with intellectual disability: Strengths and limitations in specific motor and process skills. *Australian Occupational Therapy Journal*, 50, 195–204.
45. Stauffer, L. M., Fisher, A. G., & Duran, L. (2000). ADL performance of black Americans and white Americans on the assessment of motor and process skills. *American Journal of Occupational Therapy*, 54, 607–613.
46. Lerdal, A. (2009). *Energy and fatigue in individuals with multiple sclerosis—a multi-method approach*. Saarbrücken: VDM Verlag D. Müller.
47. Barroso, J., & Lynn, M. R. (2002). Psychometric properties of the HIV-Related Fatigue Scale. *Journal of the Association of Nurses in AIDS Care*, 13, 66–75.
48. Voss, J. G., Dodd, M., Portillo, C., & Holzemer, W. (2006). Theories of fatigue: application in HIV/AIDS. *Journal of the Association of Nurses in AIDS Care*, 17, 37–50.
49. Lerdal, A. (2002). A theoretical extension of the concept of energy through an empirical study. *Scandinavian Journal of Caring Science*, 16, 197–206.
50. Mills, R., Young, C., Nicholas, R., Pallant, J., & Tennant, A. (2009). Rasch analysis of the fatigue severity scale in multiple sclerosis. *Multiple Sclerosis*, 15, 81–87.
51. Lerdal, A., Johansson, S., Kottorp, A., & von Koch, L. (2010). Psychometric properties of the fatigue severity scale: Rasch analyses of responses in a Norwegian and a Swedish MS cohort. *Multiple Sclerosis*, 16, 733–741.
52. Lee, K. A., Gay, C., Portillo, C. J., Coggins, T., Davis, H., Pullinger, C. R., et al. (2009). Symptom experience in HIV-infected adults: A function of demographic and clinical characteristics. *Journal of Pain and Symptom Management*, 38, 882–893.
53. Silverberg, M. J., Jacobson, L. P., French, A. L., Witt, M. D., & Gange, S. J. (2009). Age and racial/ethnic differences in the prevalence of reported symptoms in human immunodeficiency virus-infected persons on antiretroviral therapy. *Journal of Pain and Symptom Management*, 38, 197–207.