The comparability of Functional Assessment of Chronic Illness Therapy-Fatigue scores between cancer and systemic sclerosis

Adina Coroiu1, Linda Kwakkenbos2,3,4*, Brooke Levis3,5, Marie Hudson3,6, Murray Baron3,6, David Cella7, Brett D. Thombs1-6,8,9 and the Canadian Scleroderma Research Group9

Departments of 1Educational and Counselling Psychology and 2Psychiatry, McGill University, Montreal, Quebec, Canada; 3Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; 4Behavioural Science Institute, Clinical Psychology, Radboud University, Nijmegen, the Netherlands; Departments of 5Epidemiology, Biostatistics, and Occupational Health and 6Medicine, McGill University, Montreal, Quebec, Canada; 7Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; 8Department of Psychology, McGill University, Montreal, Quebec, Canada; 9Canadian Scleroderma Research Group recruiting rheumatologists: M. Baron, Montreal, Quebec; M. Hudson, Montreal, Quebec; G. Gyger, Montreal, Quebec; J. Pope, London, Ontario; M. Larché, Hamilton, Ontario; N. Khalidi, Hamilton, Ontario; A. Masetto, Sherbrooke, Quebec; E. Sutton, Halifax, Nova Scotia; D. Robinson, Winnipeg, Manitoba; T.S. Rodriguez-Reyna, Mexico City, Mexico; D. Smith, Ottawa, Ontario; C. Thorne, Newmarket, Ontario; P.R. Fortin, Quebec, Quebec; M. Fritzler, Mitogen Advanced Diagnostics Laboratory, Cumming School of Medicine, Calgary, Alberta

*Corresponding author: Linda Kwakkenbos, PhD; Jewish General Hospital; 4333 Cote Ste Catherine Road; Montreal, Quebec H3T 1E4; Tel: (514) 340-8222 ext. 8578; Email: kwakkenbosL@gmail.com
Acknowledgements: Not applicable.

Compliance with ethical standards

Disclosure of potential conflicts of interest: Dr. Cella developed the FACIT-F, but received no compensation for its use in these studies. Ms. Coroiu, Dr. Kwakkenbos, Ms. Levis, Dr. Marie Hudson, Dr. Baron, and Dr. Thombs declare that they have no conflict of interest.

Funding: The Canadian Scleroderma Research Group (CSRG) was funded by the CIHR (grant #FRN 83518), the Scleroderma Society of Canada and its provincial chapters, Scleroderma Society of Ontario, Sclérodermie Québec, Cure Scleroderma Foundation, INOVA Diagnostics Inc. (San Diego, CA), Dr. Fooke Laboratorien GmbH (Neuss, Germany), Euroimmun (Lubeck, Germany), Mikrogen GmbH (Neuried, Germany), FRQS, the Canadian Arthritis Network, and the Lady Davis Institute for Medical Research of the Jewish General Hospital, Montreal, QC. The CSRG has also received educational grants from Pfizer and Actelion pharmaceuticals. Ms. Coroiu was supported by a CIHR Doctoral Research Award. Dr. Kwakkenbos was supported by a CIHR Banting Postdoctoral Fellowship. Ms. Levis was supported by a CIHR Doctoral Research Award. Dr. Thombs was supported by an Investigator Salary Award from the Arthritis Society.

Authors’ contributions: Ms. Coroiu, Dr. Kwakkenbos, and Dr. Thombs made substantial contributions to the conception and design, analysis and interpretation of data, and were involved in the drafting of the final manuscript. Ms. Levis made substantial contributions to the data coding and provided feedback on the final draft of the manuscript. Dr. Cella, Dr. Hudson, and Dr. Baron made substantial contributions to the acquisition of data and provided feedback on the final draft of the manuscript. All authors read and approved the final manuscript. Dr. Kwakkenbos is the overall guarantor.
Research Involving Human Participants: “All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.”

Informed Consent: “Informed consent was obtained from all individual participants included in the study.”
Abstract

**Purpose:** The Functional Assessment of Chronic Illness Therapy- Fatigue (FACIT-F) is commonly used to assess fatigue across diseases. The degree to which the FACIT-F demonstrates measurement equivalence across disease groups, however, is not known. The purpose of this study was to assess differential item functioning (DIF) of FACIT-F items between patients with cancer and systemic sclerosis (SSc, or scleroderma).

**Methods:** Secondary analysis of FACIT-F data from cancer and SSc patients. Confirmatory factor analysis was used to assess the factor structure of the FACIT-F in cancer and SSc patients. The Multiple-Indicator Multiple-Cause model was utilized to assess DIF, comparing responses from cancer and SSc patients.

**Results:** A unidimensional factor structure for the FACIT-F was demonstrated with the cancer (n=1141), SSc (n=1186), and combined samples. Statistically significant, but small-magnitude, DIF was found for four items. Compared to cancer patients with the same level of fatigue, SSc patients had lower scores (more fatigue) for item 2 (*bodily weakness*), 7 (*energy*), and 8 (*ability to perform daily activities*); and higher scores (less fatigue) for item 9 (*need to sleep throughout the day*). For the entire scale, SSc patients had 0.47 SD lower FACIT-F latent factor scores (more fatigue) than cancer patients. After correcting for DIF, there was a change of only 0.03 SD in this difference (0.44 SD lower).

**Conclusions:** Although statistically significant DIF was detected for four FACIT-F items, the magnitude was small and the effect on fatigue latent scores was minimal. Thus, FACIT-F scores can be used equivalently in cancer and SSc.

**Keywords:** FACIT-F; systemic sclerosis; cancer; measurement equivalence; differential item functioning
INTRODUCTION

Chronic fatigue from medical illness involves ongoing exhaustion (e.g., weakness, lack of energy, and tiredness) that is disproportionate to exertion and not relieved by rest [1]. Debilitating levels of fatigue are common and a major factor in decreased health-related quality of life (HRQL) in many chronic diseases [2]. Despite this, chronic illness-related fatigue is under researched, poorly understood, and often overlooked by clinicians [2]. Contributing to this, there is a lack of uniformity of assessment methods, with more than 250 measurement scales and single-question inquiries available to assess fatigue in chronic disease [3]. One reason is that untested assumptions about differences in fatigue presentation across diseases has led to the development of disease-specific measurement scales [4]. Common methods for assessing fatigue, which could be applied broadly across medical populations, would facilitate comparison of fatigue presentation, research on etiology and treatment, and would contribute to the standardizing of communication about fatigue.

The Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) was originally developed to assess fatigue in people with cancer [5], but has been used in studies of patients with many different diseases and has been translated into more than 50 languages. Whether or not the FACIT-F is measurement equivalent across diseases, however, has not been demonstrated. Assessments of measurement equivalence are done to determine whether patients with different characteristics, but who have the same level of the construct being measured, respond similarly to items assessing the construct and that item responses are not substantively influenced by other patient characteristics. Assessment tools that are measurement equivalent produce unbiased, comparable estimates of the construct being measured across groups [6]. Differential item functioning (DIF), on the other hand, is said to occur when patients with different diseases, but with similar levels of a construct, such as fatigue, score differently on items assessing the construct. DIF between disease groups could occur, for example, if items are more relevant to patients in one of the groups or if items are perceived...
or interpreted differently across groups, and if these differences lead to different items responses even when levels of the outcome being measured are similar [7].

We do not know of any previous studies that have evaluated the degree to which the FACIT-F is measurement equivalent across diseases. SSc is a chronic connective tissue disorder characterized by thickening and fibrosis of the skin, internal organ involvement, poor HRQL, and significant morbidity and mortality [8-10]. Similar to cancer [11-14], SSc patients report that fatigue is common and that it impacts HRQL as much or more than any other symptom [15-17]. A previous study assessed the measurement equivalence of cancer-related fatigue case-definition criteria and found comparable rates of fatigue between SSc (35.1%) and breast cancer (37.8%) patients; however, DIF analysis found that two items, i.e., “trouble concentrating” and “short-term memory problems” were more frequently endorsed in cancer compared to SSc [18]. The objective of the present study was to evaluate the measurement equivalence of the FACIT-F scale in patients with cancer and SSc.

**METHOD**

**Patients and Procedures**

Patients with cancer were drawn from three separate samples recruited in the United States. We conducted secondary analyses with data that have already been collected and combined three cancer databases to yield a larger cancer sample size comparable to that of the scleroderma sample. SSc patients were drawn from a large pan-Canadian registry. Patients from both datasets who provided complete item data on the FACIT-F were included in analyses. Ethical approval was obtained from local ethics boards at the enrolling sites for all samples, and all patients provided informed consent.

*Cancer - Bone Marrow Transplant (BMT) Sample* [19]. Patients who had undergone allogeneic and autologous BMT at least one year prior were recruited through the International Bone Marrow Transplant Registry and the Autologous Blood and Marrow Transplant Registry between
March 2000 and September 2002. A combination of self-report questionnaires, including the FACIT-F, were administered in a cross-sectional design.

**Cancer - CaPS-MID Sample [20]**. Patients with various cancer diagnoses (e.g., bone, brain, breast, colorectal, gynaecological, head/neck, leukemia, liver, lung) were recruited from Evanston Northwestern Healthcare and Northwestern University, Illinois from 2005 to 2006. Self-report questionnaires, including the FACIT-F, were completed on two separate occasions; however, for the current analysis, only FACIT-F data collected at the first time point were used.

**Cancer – Goal Interference (GI) Sample [21]**. Patients receiving chemotherapy treatment for a first diagnosis of a solid tumour or lymphoma, and who had been diagnosed less than two months before enrolment, were recruited from Evanston Northwestern Healthcare and Northwestern University, Illinois. Recruitment took place between May 2002 and February 2005. Questionnaires, including the FACIT-F, were conducted at baseline and two follow-up time points. For the current study, only FACIT-F data collected at the first time point were used.

**SSc Sample**. Patients recruited from the Canadian Scleroderma Research Group (CSRG) Registry completed the FACIT-F between March 2008 and December 2013. Eligibility criteria for the Registry included having a diagnosis of SSc confirmed by a CSRG rheumatologist, being at least 18 years of age, and being fluent in English or French. Patients in the Registry, who are recruited across 15 centres located across Canada, undergo extensive physical evaluations at annual visits and complete a series of self-report questionnaires in their preferred language (English or French). For patients who completed the FACIT-F on multiple occasions, the last available visit with complete FACIT-F data was used.

**Measures**

**Demographics and Disease Characteristics**. Demographic variables collected for all samples included age, sex, educational level, marital status, and current employment status. Medical variables
included type of diagnosis, time since diagnosis and/or treatment, and disease severity (i.e., cancer stage or SSc subtype). In the cancer samples, medical variables were collected by the research team directly from the patients’ medical charts and from self-report questionnaires. In the SSc sample, a CSRG rheumatologist provided time since diagnosis and a classification of limited versus diffuse SSc. Limited SSc was defined as skin involvement distal to the elbows and knees only, whereas diffuse SSc was defined as skin involvement proximal to the elbows and knees, and/or the trunk [22].

*Functional Assessment of Chronic Illness Therapy- Fatigue (FACIT-F) [5].* The FACIT-F comprises 13 items that assess tiredness, weakness, and difficulty conducting everyday activities due to fatigue in the past 7 days. Items are scored on a 5-point Likert-type scale ranging from 0 (*not at all*) to 4 (*very much*). All items except items 7 (“I have energy”) and 8 (“I am able to do my usual activities”) are reverse-scored before being summed to obtain a total score (range 0-52), with higher scores reflecting lower levels of fatigue. The FACIT-F has been shown to have excellent internal consistency (Cronbach’s alpha > 0.90) and strong concurrent, divergent, and predictive validity in several samples of patients with cancer and other chronic health problems [23]. The FACIT-F was completed in English by cancer patients and in English or French by SSc patients. A previous comparison of English and French FACIT-F scores in SSc patients showed that the measure has equivalent score metrics in these languages [24].

**Data Analyses**

Demographics characteristics (age, sex, marital status) were compared between cancer and SSc patients, using the chi-square statistic for categorical variables and t-tests for continuous variables. The factor structure of the FACIT-F was initially assessed in each sample separately using confirmatory factor analysis (CFA). Ideally, for DIF assessment, the most parsimonious model with a reasonable fit is used. The FACIT-F has been shown to be a unifactorial measure across various samples [25]. Thus, a single-dimensional CFA model was fit in order to determine whether this structure could be reasonably
used in the DIF analysis. Item responses for the FACIT-F were ordinal Likert data and were therefore modeled using the weighted least squares estimator with a diagonal weight matrix, robust standard errors, and a mean- and variance-adjusted chi-square statistic with delta parameterization [26]. In addition to the chi-square test, which is highly sensitive to sample size and can lead to erroneous rejection of models with reasonable fit [27], the Tucker-Lewis Index (TLI) [28], the Comparative Fit Index (CFI) [29] and the Root Mean Square Error of Approximation (RMSEA) [30] were used to assess the model fit. Good fitting models are indicated by a TLI and CFI ≥ 0.95 and RMSEA ≤ 0.06 [31], although a CFI and TLI of .90 or above [32] and a RMSEA of .08 or less [33] are often regarded as indicators of an adequate model fit. Modification indices were used to identify pairs of items for which the model fit would improve if the error estimates were free to covary and for which there appeared to be theoretically justifiable shared method effects (e.g., similar content) [34]. After establishing the factor structure in each sample, the 1-factor model was fit with both samples combined.

In order to determine whether any of the FACIT-F items exhibited DIF in the cancer versus SSc sample, the Multiple-Indicator Multiple-Cause (MIMIC) model was utilized. MIMIC models for DIF assessment are based on structural equation models, in which the grouping variable (cancer versus SSc) is added to the basic model as an observed variable. The base MIMIC model consists of the factor model, to which the additional direct effect of group on the latent factor is added. This serves to control for group differences on the level of the latent factor. An important strength of the MIMIC model is that it allows for adjustment of covariates that may differ between comparison groups, by adding a direct effect of these variables on the latent factors. We used age and sex as covariates in the analyses. Thus, the base model for testing DIF included the direct effect of group (cancer versus SSc) and covariates (age, sex) on the latent fatigue factor.

Each FACIT-F item was then regressed separately on the group variable to assess for potential
DIF. DIF was determined to be present for items with a statistically significant association of group with the item, controlling for differences in the overall level of the fatigue latent factor between groups. If DIF for one of the items was identified, the association between the group variable and the item with the largest magnitude of DIF was added to the model. This procedure was repeated until none of the remaining items showed statistically significant DIF. At each step of the comparison, Hommels’ correction for multiple testing was applied [35]. Once all of the items with significant DIF were identified, the magnitude of DIF items collectively was evaluated by comparing the difference on the latent factor between groups in the base model and in the model controlling for DIF. The magnitude of this difference was interpreted following Cohen’s effect sizes, with ≤ 0.20 SD indicating small, 0.50 SD = moderate and 0.80 SD = large differences [36-38]. As a sensitivity analysis, the MIMIC model was repeated, excluding all items with statistically significant DIF. In additional sensitivity analyses, we conducted DIF analyses for each cancer, separately, compared to the SSc sample. CFA and DIF analyses were conducted using MPlus 7 [26] and all other analyses were conducted using IBM SPSS Statistics 22 (Chicago, IL).

RESULTS

Sample Characteristics

There were 1141 cancer patients and 1186 SSc patients included. In the BMT sample (n=637), time since treatment ranged from 2 - 23 years. In the GI (n=207) and CaPS-MID (n=297) samples, time since diagnosis ranged from 0 - 2 months and 0 - 29 years, respectively. Among SSc patients, time since diagnosis ranged from 0 – 50 years. Further demographic and disease characteristics are displayed in Table 1. Cancer patients were significantly younger than SSc patients (p < .001), and there were fewer female cancer patients (p < .001). Marital status was similar across the two samples.
Confirmatory Factor Analysis

A unifactorial factor model was initially assessed in the cancer ($\chi^2(65) = 1800.08, p < 0.001; \text{CFI} = .96; \text{TLI} = .96; \text{RMSEA} = 0.15$) and SSc samples ($\chi^2(65) = 1514.36, p < 0.001; \text{CFI} = .98; \text{TLI} = .98; \text{RMSEA} = 0.14$) separately. Upon examining the modification indices, three pairs of error terms pertaining to items 5 and 6, 4 and 1, and 7 and 8, respectively, were allowed to freely covary. Items 5 (“trouble starting things”) and 6 (“trouble finishing things”) and 1 (“feeling fatigued”) and 4 (“feeling tired”), respectively, had similar content, whereas items 7 (“have energy”) and 8 (“able to do my usual activities”) were the only two FACIT-F items that were phrased positively. Freeing these three error covariances resulted in reasonably adequate model fit in both samples (cancer: $\chi^2(62) = 878.06, p < 0.001; \text{CFI} = .98; \text{TLI} = .98; \text{RMSEA} = 0.11$; SSc: $\chi^2(62) = 876.25, p < 0.001; \text{CFI} = .99; \text{TLI} = .99; \text{RMSEA} = 0.11$) and the combined sample ($\chi^2(62) = 1661.37, p < 0.001; \text{CFI} = .99; \text{TLI} = .98; \text{RMSEA} = 0.11$).

DIF Analysis

MIMIC Base Model. The single-factor structure was fit to the combined sample of cancer and SSc patients (N=2327), adding the group (cancer and SSc) and covariate (age, sex) variables to the model. This model had adequate fit ($\chi^2(98) = 1881.55, p < 0.001, \text{CFI} = .99, \text{TLI} = .98, \text{RMSEA} = 0.09$). Prior to accounting for possible DIF, SSc patients had 0.47 SD (95% confidence interval [CI] - 0.57 to -0.38; $p < .001$) lower FACIT-F latent factor scores (more fatigue) than cancer patients. See Table 2.

DIF Assessment. Four items showed statistically significant DIF: item 2 ($z = -5.69, p < 0.001$), item 7 ($z = -6.15, p < 0.001$), item 8 ($z = -2.83, p < 0.001$), and item 9 ($z = 3.77, p < 0.001$). Items 2 (“I feel weak all over”), 7 (“I have energy”), and 8 (“I am able to do my usual activities”) were rated at lower levels (more fatigue) in the SSc sample, whereas item 9 (“I need to sleep during the day”) was rated at a higher level (less fatigue) in the SSc sample, compared to the cancer sample. After
controlling for the cumulative DIF of these 4 items, SSc patients had 0.44 SD (95% CI -0.53 to -0.35; \( p < .001 \)) lower FACIT-F latent factor scores (more fatigue) than cancer patients. Thus, compared to the base model, the DIF corrected model showed a decrease of 0.03 SD in the effect of group (cancer versus SSc) on the overall latent fatigue scores (small magnitude). The factor loadings of the DIF corrected model as well as the effects of the DIF items and of group (cancer and SSc), respectively, on the fatigue latent factor are displayed in Table 2. The results of the sensitivity analysis, in which the 4 DIF items were omitted from the model, were virtually identical to the 13-item DIF corrected model, and the difference between groups on the latent factor was -0.44 SD (95% CI -0.53 to -0.35; \( p < .001 \)).

DIF analyses that compared each cancer separately with the SSc sample also found that differences in fatigue between the SSc and cancer patients did not differ substantively between the base model and the DIF corrected model (GI and SSc: -0.54 versus -0.45 SD; CaPS-MID and SSc: -0.54 SD versus -0.51 SD; BMT and SSc: -0.56 versus -0.57 SD).

**DISCUSSION**

The main results of this study were that (1) four items of the FACIT-F showed statistically significant, but small magnitude, DIF when comparing item responses of cancer and SSc patients, but that (2) the cumulative impact of item-level DIF on fatigue latent scores was negligible, suggesting that FACIT-F scores can be treated equivalently in cancer and SSc patients.

One other study [18] compared the measurement of fatigue in patients with breast cancer and SSc using case-definition criteria. That study found that cancer patients reported more cognitive difficulties, reflected in items on concentration and short-term memory difficulties, than SSc patients with the same level of fatigue, and that this would likely influence the overall assessment of fatigue [18]. The FACIT-F does not assess cognitive symptoms and we found that although there were some statistically significant differences on individual items, these differences were small and did not influence the overall measurement of fatigue. Statistically significant findings can arise from
magnitudes of importance or due to using large sample sizes. Our sample included over 2000 patients, and statistically significant differences in DIF analyses were not of magnitude large enough to interpret. Furthermore, assessment of potential DIF involves a large number of analyses, which can lead to false positive findings and could explain some of the statistically significant item DIF that we identified in the present study. Given that all differences were deemed minimal, as per Cohen’s guidelines, the scoring metric of the FACIT-F can be reasonably treated as equivalent in cancer and SSc samples. Future research should assess the measurement invariance of FACIT-F across other diseases as well as in cancer survivors at different points in the illness trajectory (e.g., during and immediately post treatment versus throughout survivorship) to better understand fatigue in the context of chronic illness.

There are limitations that should be considered when interpreting the results of this study. First, in all analyses the RMSEA coefficient exceeded the common threshold of .06. This is consistent with results from previous CFA’s conducted with the FACIT-F [23, 25], however, and is not particularly problematic given the strong CFI and TLI indices, which taken together support the unidimensionality of the measure. Second, different recruitment strategies were used in the two samples. The SSc sample was recruited from 15 centres located across Canada, whereas the cancer patients were recruited in hospitals in the United States. Lastly, a potential disadvantage of the MIMIC model, as compared to other models that test for measurement invariance, is that it does not test for non-uniform DIF, which would identify whether the DIF varied at different levels of the construct being assessed. The MIMIC model, however, does allow for the adjustment of relevant covariates, such as demographic variables that were different between cancer and SSc patients.

In sum, this study found that four items of the FACIT-F had statistically significant DIF in cancer versus SSc samples. However, the magnitude of the four items’ cumulative impact on the fatigue latent scores was negligible. Thus, this study provided empirical support for the measurement
equivalence of the FACIT-F in cancer and SSc. Further research is needed to establish the measurement equivalence of the FACIT-F across other chronic diseases.

Word count: 2912
REFERENCES


Table 1

Demographic and Disease Characteristics of the Cancer and SSc Samples

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cancer</th>
<th>SSc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n %</td>
<td>719 (63)</td>
<td>1033 (87)</td>
</tr>
<tr>
<td>Age in years, M (SD), Range</td>
<td>53.0 (12.4), 18-90</td>
<td>59.0 (11.8), 18-88</td>
</tr>
<tr>
<td>Education level, n (%) &gt; 12 years</td>
<td>659 (70)</td>
<td>606 (51)</td>
</tr>
<tr>
<td>Married or living as married, n (%)</td>
<td>708 (72)</td>
<td>809 (68)</td>
</tr>
<tr>
<td>Type of cancer diagnosis, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>250 (26)</td>
<td></td>
</tr>
<tr>
<td>Colorectal</td>
<td>54 (6)</td>
<td></td>
</tr>
<tr>
<td>Gynaecological</td>
<td>31 (3)</td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>32 (3)</td>
<td></td>
</tr>
<tr>
<td>Leukemia</td>
<td>366 (38)</td>
<td></td>
</tr>
<tr>
<td>Hodgkin’s or non-Hodgkin’s lymphoma</td>
<td>150 (16)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>76 (8)</td>
<td></td>
</tr>
<tr>
<td>Severity of diagnosis, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced stage (III or IV)</td>
<td>417 (79)</td>
<td>597 (51)</td>
</tr>
<tr>
<td>Limited disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FACIT-F score, M (SD), Range</td>
<td>37.0 (11.3), 2-52</td>
<td>31.7 (12.5), 0-52</td>
</tr>
</tbody>
</table>

Note. SSc=systemic sclerosis; M=mean; SD=standard deviation

a n=637 for BMT sample, n=297 for CaPS-MID sample, n=207 for GI sample.
b n=941; c n=990; d n=1180; e n=959; f n=526, for the CaPS-MID and GI samples only; g n=1167.
Table 2

*Factor Loadings for the FACIT-F in Cancer and SSc Patients and the Impact on the Overall Estimates of Fatigue Latent Factor Scores*

<table>
<thead>
<tr>
<th>Item</th>
<th>Base model(^a)</th>
<th>DIF corrected model(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor loading</td>
<td>95% CI</td>
</tr>
<tr>
<td>1. I feel fatigued</td>
<td>0.86</td>
<td>[0.85, 0.87]</td>
</tr>
<tr>
<td>2. I feel weak all over</td>
<td>0.87</td>
<td>[0.85, 0.88]</td>
</tr>
<tr>
<td>3. I feel listless (“washed out”)</td>
<td>0.91</td>
<td>[0.90, 0.92]</td>
</tr>
<tr>
<td>4. I feel tired</td>
<td>0.90</td>
<td>[0.89, 0.91]</td>
</tr>
<tr>
<td>5. I have trouble starting things because I am tired</td>
<td>0.89</td>
<td>[0.88, 0.90]</td>
</tr>
<tr>
<td>6. I have trouble finishing things because I am tired</td>
<td>0.87</td>
<td>[0.85, 0.88]</td>
</tr>
<tr>
<td>7. I have energy</td>
<td>0.75</td>
<td>[0.73, 0.77]</td>
</tr>
<tr>
<td>8. I am able to do my usual activities</td>
<td>0.68</td>
<td>[0.66, 0.71]</td>
</tr>
<tr>
<td>9. I need to sleep during the day</td>
<td>0.68</td>
<td>[0.65, 0.70]</td>
</tr>
<tr>
<td>10. I am too tired to eat</td>
<td>0.74</td>
<td>[0.71, 0.77]</td>
</tr>
<tr>
<td>11. I need help doing my usual activities</td>
<td>0.73</td>
<td>[0.70, 0.75]</td>
</tr>
<tr>
<td>12. I am frustrated by being too tired to do the things I want to do</td>
<td>0.88</td>
<td>[0.87, 0.89]</td>
</tr>
<tr>
<td>13. I have to limit my social activity because I am tired</td>
<td>0.88</td>
<td>[0.87, 0.89]</td>
</tr>
</tbody>
</table>
Direct effects on items attributable to group (Cancer versus SSc):

<table>
<thead>
<tr>
<th>Item</th>
<th>Direct Effect</th>
<th>CI Lower</th>
<th>CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 2</td>
<td>-0.18</td>
<td>[-0.24, -0.12]</td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td>-0.22</td>
<td>[-0.29, -0.15]</td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td>-0.11</td>
<td>[-0.18, -0.03]</td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td>0.14</td>
<td>[0.07, 0.22]</td>
<td></td>
</tr>
</tbody>
</table>

Structural effect of group (Cancer versus SSc) on the latent factor | -0.47 | [-0.57, -0.38] | -0.44 | [-0.53, -0.35] |

*Note. FACIT-F = FACIT Fatigue Scale; SSc = systemic sclerosis; DIF = Differential Item Functioning; CI = confidence interval

aNot corrected for DIF. bCorrected for DIF for items 2, 7, 8, and 9.