

Measuring fatigue in SSc: a comparison of the Short Form-36 Vitality subscale and Functional Assessment of Chronic Illness Therapy–Fatigue scale

Daphna Harel¹, Brett D. Thombs^{2,3,4,5,6}, Marie Hudson^{3,6}, Murray Baron^{3,6} and Russell Steele^{1,6}, on behalf of Canadian Scleroderma Research Group*

Abstract

Objective. Fatigue is a common and important problem in SSc. No studies, however, have compared the properties of fatigue measures in SSc. The objective of this study was to compare the performances of the Short Form-36 (SF-36) Vitality subscale and Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT) in SSc.

Methods. Cross-sectional, multi-centre study of Canadian Scleroderma Research Group Registry patients. The associations of the two instruments with other patient-reported outcome measures, as well as physician- and patient-rated disease variables were compared. Item response theory models were used to compare the degree to which items and the total scores of each measure effectively covered the full spectrum of fatigue levels.

Results. There were 348 patients (297 women, 85%) in the study. The instruments correlated at $r=0.65$ with each other. The FACIT tended to correlate slightly higher than the SF-36 Vitality subscale with physician- and patient-rated disease variables and patient-reported physical function and disability, whereas the SF-36 Vitality subscale correlated minimally higher with mental health measures. The FACIT had markedly better discrimination across the range of fatigue, particularly at average to high fatigue levels, whereas the SF-36 Vitality subscale discriminated well only among patients in the low to average range.

Conclusion. The FACIT discriminates better than the SF-36 Vitality subscale at average to high ranges of fatigue, which is common in SSc, suggesting that it is preferred for measuring fatigue in SSc.

Key words: SSc, scleroderma, fatigue, psychometrics, FACIT, SF-36 Vitality.

Introduction

SSc is a chronic multi-system disorder characterized by microvascular disease, disturbance in fibroblast function and immune system activation, culminating in fibrosis of

skin and internal organs [1]. SSc affects mostly women [2] and is associated with significant disability [3], high cost [4] and increased mortality [5]. Common morbidities include disfiguring skin thickening, finger ulcers, joint contractures, pulmonary hypertension, interstitial lung disease, arthritis, chronic dyspepsia, malabsorption and malnutrition and renal failure. Patients have high rates of clinically significant symptoms of depression, even compared with patients with other acute and chronic conditions [6]. SSc is associated with substantially reduced health-related quality of life (HRQoL) [7].

Persistent fatigue, defined as ongoing exhaustion that is disproportionate to exertion and not alleviated by rest, reduces HRQoL for patients with many chronic medical illnesses [8]. A recent study reported that fatigue was present at least some of the time among 89% of > 400

¹Department of Mathematics and Statistics, ²Department of Psychiatry, ³Department of Medicine, ⁴Department of Epidemiology, Biostatistics and Occupational Health, ⁵Department of School of Nursing, McGill University, and ⁶Department of Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

Submitted 4 January 2012; revised version accepted 26 June 2012.

Correspondence to: Brett D. Thombs, Jewish General Hospital, 4333 Cote Ste Catherine Road, Montreal, Quebec H3T 1E4, Canada. E-mail: brett.thombs@mcgill.ca

*See supplementary data for a list of the CSRG Recruiting Rheumatologists, available at *Rheumatology* Online.

Canadian SSc patients and had a moderate to severe impact on the ability to carry out daily activities for 72% [9]. Similarly, a Dutch study found that fatigue was reported to be a bothersome symptom for 92% of 123 SSc patients [10]. Fatigue ratings by SSc patients are similar to those of patients with other rheumatic diseases and cancer patients currently undergoing treatment, and worse than in the general population and among cancer patients in remission [11]. Fatigue in SSc is associated with reduced capacity to carry out daily activities, work disability and impaired physical function, even after controlling for education level, disease subtype, pain, sleep quality and depressive symptoms [12–15]. Factors that have been associated with fatigue in SSc include breathing and gastrointestinal problems, pain and symptoms of depression [16]. An important limitation of existing studies on fatigue in SSc is that they have used single-item ratings [9, 10] or measurement scales that have not been validated for patients with SSc, including visual analogue scales [13, 14], the Vitality subscale of the SF-36 [12, 16, 17] and the Multidimensional Assessment of Fatigue (MAF) scale [15, 18]. For example, Ibn Yacoub *et al.* [19] found a strong negative correlation between the SF-36 Vitality subscale and the MAF, but Sandqvist *et al.* [20] and Rannou *et al.* [21] came to the opposite conclusion.

The SF-36 Vitality subscale and the Functional Assessment of Chronic Illness Therapy (FACIT) scale are widely used to measure fatigue across chronic disease groups, including the rheumatic diseases. The SF-36 Vitality subscale consists of four items and was designed to assess fatigue in healthy respondents from the general population [17]. The 13-item FACIT, on the other hand, was developed to assess fatigue among patients with chronic medical disease [22]. Cella *et al.* [23] compared the psychometric properties of the SF-36 Vitality subscale and FACIT in >600 RA patients. Standard indices of reliability and validity were good for both measures. However, the SF-36 Vitality subscale did not adequately differentiate among patients with moderate versus severe fatigue because patients with at least moderate fatigue tended to score at high levels on the SF-36 Vitality subscale items. The FACIT, on the other hand, discriminated well across the range of fatigue.

The main objective of this study was to compare the measurement properties of the SF-36 Vitality subscale and the FACIT in patients with SSc. To do this, we (i) compared the associations of the two instruments with other physician- and patient-reported outcome measures and (ii) evaluated the extent to which individual items on the measures and their total scores discriminated between patients with different levels of fatigue.

Materials and methods

Design

Cross-sectional study of a Canadian multi-centre cohort of SSc patients.

Patients and procedure

Study participants were patients enrolled in the Canadian Scleroderma Research Group (CSRG) Registry. Patients in this registry are recruited from 15 centres across Canada. Registry patients must have a diagnosis of SSc confirmed by a rheumatologist, be ≥ 18 years of age and be fluent in English or French. Patients recruited into the Registry are assessed yearly with a standardized clinical evaluation, at which time they complete a number of self-report instruments, including the SF-36, FACIT, and measures of function, depression and pain. Registry patients were included in this study if they completed the SF-36 Vitality subscale and FACIT and had complete data for gender, disease type and disease duration recorded at an initial/baseline Registry visit that occurred between November 2009, when the FACIT was initially introduced to the Registry, and October 2010.

Measures

In addition to the SF-36 Vitality subscale and FACIT, we collected demographic and general disease variables; physician assessments of disease severity, activity and damage; patient ratings of a series of SSc symptoms from the Scleroderma HAQ (SHAQ); HAQ Disability Index (HAQ-DI) and patient-oriented outcome measures, including the SHAQ, the Center for Epidemiologic Studies Depression scale (CES-D) and SF-36 physical and mental component summary scores.

Demographic and disease variables

Disease duration was recorded by registry physicians as time since the onset of the first non-RP disease manifestation. Skin involvement was assessed using the modified Rodnan skin score [24], and patients are classified into limited and diffuse disease subsets based on Leroy's definition [25]. Patients without any skin involvement (<5%) were included with those defined as having limited disease.

Physician assessments

Physician-reported global assessments of disease activity, damage and severity [26–34] were measured using an 11-point numerical rating scale asking the rater to rate the subject's disease activity/damage/severity in the past week.

SF-36 Vitality subscale

The SF-36 version 2.0 Vitality [35] subscale consists of four Likert items. Patients rate agreement with statements related to fatigue in the past 4 weeks on a 5-point scale (1 = all the time, 5 = none of the time). Scores on the SF-36 Vitality subscale are normalized based on US population data ($M=50$, $SD=10$) [17]. In the tables, we report the scores consistent with normalized scoring. However, in order for the scoring direction to be consistent with the FACIT in analyses, the response ordering of the SF-36 Vitality subscale items was coded so that higher ratings indicated greater fatigue. The SF-36 Vitality subscale has been shown to possess reasonably good internal

consistency (Cronbach's $\alpha = 0.80$) and good discriminative validity in the Canadian population when administered in both English and Canadian French [36].

FACIT

The FACIT consists of 13 Likert items, measured on a 5-point scale (0 = not at all, 4 = very much) in which patients rate statements about fatigue in the past 7 days. Scores are summed to obtain a total score (range 0–52). The FACIT has been shown to have excellent internal consistency (Cronbach's $\alpha > 0.90$) and very good construct and divergent validity in several patient populations [37, 38].

Other patient-reported instruments

In addition to the physician-reported assessments, we used patient-reported measures to describe the cohort and to assess their correlations with the SF-36 Vitality subscale and the FACIT. The instruments used in our study include modified SF-36 physical component (PCS) and mental component (MCS) summaries [39], patient ratings of a series of SSc symptoms from the SHAQ [40–44], and the CES-D [45–47]. Further details about these instruments can be found in the [supplementary data](#), available at *Rheumatology* Online.

Statistical analysis

To assess the association between the total scores of the two instruments, we computed the correlation of the SF-36 Vitality subscale and FACIT total scores. In addition, correlations were computed to compare associations of the SF-36 Vitality subscale and the FACIT with physician- and patient-rated outcome measures. All correlations, with 95% CIs, were calculated for 1000 bootstrap samples using the non-parametric Kendall's tau test. Bias-corrected CIs for the differences between correlations of the SF-36 Vitality subscale and FACIT with the measures were also computed [48].

We used item response theory (IRT) methods to examine the degree to which items and the total scores of the SF-36 Vitality subscale and FACIT differentiated between patients' fatigue levels. We validated two of the most important assumptions of the model, unidimensionality of the latent construct and stochastic ordering. Unidimensionality was assessed in two different ways. First, we fitted a confirmatory factor analysis model with a separate latent construct for each questionnaire [49], as well as a model that fitted both to a single latent construct. Models were compared on the basis of their Bayesian information criterion (BIC) values [50, 51], as well as root mean squared error of approximation (RMSEA) [52]. As per the guidelines of Raftery and Kass [50], the model with the lowest BIC value was chosen, with a difference of 10 in BIC values considered to be large. Furthermore, the model with the lowest RMSEA was selected. Second, we fitted a partial credit IRT model (PCM) [53] to the SF-36 Vitality subscale data and the FACIT data separately, which assumes two independent latent constructs. We compared the single latent trait model with the separate latent trait model using the BIC. Stochastic ordering

was assessed by checking manifest monotonicity for each item by grouping patients based on their scores on the rest of the instrument [54]. Major violations to the assumption of stochastic ordering were assessed using the guidelines of Van de Ark, with a violation > 0.3 considered relevant [55].

A PCM was fitted to the combined data for both instruments. Fatigue was assumed to have a standard normal distribution in the population. For each item, the PCM model estimates the level of fatigue at which a patient would choose one score category instead of the category below it, this threshold parameter reflects the fatigue level at which patients would be expected to endorse a more severe item category. Because both the SF-36 Vitality subscale and FACIT items have five levels, there were four fatigue thresholds between response levels for each item of both scales. These item-specific threshold parameters were then used to compare the range of fatigue measured by each item, as well as to assess how patients with varying levels of fatigue would be expected to respond to a given item. As done in Cella *et al.* [23], to facilitate comparison between the sets of items, the four threshold parameters for each item are averaged to create a single representative value.

The fatigue level for each pattern of item response can be estimated by using the parameters of the PCM and an application of Bayes Theorem. These can be interpreted as the expected value for the latent fatigue construct for an individual given the categories that they responded to for each of the items. We also reviewed the most likely scores that corresponded to a wide range of latent fatigue values to determine how well each instrument's total score differentiated patients at varying levels of fatigue.

Although *a priori* *P*-values < 0.05 were considered statistically significant, we applied a post hoc Bonferroni correction factor for each of the 26 comparisons in Table 2 (< 0.002). All analyses were done using the R statistical package [56], version 2.12.0. All 95% CIs were obtained using a non-parametric bootstrap via the boot package [57, 58]. The test of stochastic ordering was conducted using the mokken package [55]. All PCMs were fitted using the ltm package [59].

This study was approved by the McGill University Institutional Review Board, and the data collection protocol was approved by the ethics committees of all study sites. Informed consent to participate was obtained from all study participants.

Results

The baseline characteristics for patients included in this study ($N = 348$) are shown in Table 1. Approximately 85% of patients were female, and the mean age was 56 years. Mean time since onset of the first non-RP symptoms was just over 11 years, and 35% of patients were classified as diffuse. The mean score on the SF-36 Vitality subscale was 44.8 (s.d. = 11.2), or half a s.d. below the US population norm and the mean score on the FACIT was 20.0 (s.d. = 12.2).

TABLE 1 Baseline characteristics of the CSRG patients (N = 348)

Characteristic	Mean (s.d.) or n (%), N = 348	Median (IQR), N = 348
Women, n (%)	297 (85.3)	–
Disease duration since onset of first non-RP symptom, mean (s.d.)	11.1 (9.9)	8.6 (3.0–17.0)
Age, mean (s.d.), years	55.6 (11.5)	56.0 (48.0–63.0)
Disease subset, n (%)		
Diffuse	121 (34.8)	
Limited	213 (61.2)	
Sine scleroderma	14 (4.0)	
Modified Rodnan skin score, mean (s.d.)	9.1 (9.1)	6.0 (2.0–12.0)
Physician global assessments, mean (s.d.), range 0–10		
Disease severity	2.8 (2.2)	2.0 (1.0–4.0)
Disease activity	2.6 (2.2)	2.0 (1.0–3.3)
Disease damage	3.2 (2.2)	3.0 (2.0–4.0)
SF-36* mean (s.d.)		
Physical component summary score	38.6 (11.1)	38.1 (30.0–47.2)
Without the Vitality subscale	38.7 (10.9)	38.4 (30.1 – 47.3)
Mental component summary score	47.8 (11.9)	50.2 (39.9–57.5)
Without the Vitality subscale	49.1 (10.3)	51.0 (42.2–56.9)
Vitality subscale	44.8 (11.2)	45.9 (36.5–52.1)
FACIT, range 0–52, mean (s.d.)	20.0 (12.2)	19.0 (10.0–30.0)
HAQ-DI, mean (s.d.)	0.7 (0.7)	0.5 (0.0–1.3)
Patient global assessment, mean (s.d.), range 0–10		
Pain	3.4 (2.7)	3.0 (1.0–5.0)
RP	2.8 (2.8)	2.0 (0.0–5.0)
Finger ulcers	1.9 (3.0)	0.0 (0.0–3.0)
Gastrointestinal problems	1.4 (2.2)	0.0 (0.0–2.0)
Breathing	1.9 (2.5)	1.0 (0.0–3.0)
Disease severity	3.4 (2.6)	3.0 (1.0–5.0)
CES-D, mean (s.d.)	14.1 (10.6)	12.0 (5.0–21.0)

*SF-36 summary and subscale scores are normalized with a mean of 50 and s.d. of 10. Scores below 50 represent worse and above 50 represent better quality of life.

Associations of SF-36 Vitality subscale and FACIT

The correlation between the SF-36 Vitality subscale and FACIT was 0.65 (95% CI: 0.61, 0.69). Table 2 shows correlations of the two instruments with physician- and patient-rated outcome measures. The FACIT had a higher correlation than the SF-36 Vitality subscale on all physician-rated measures, although the magnitudes of the differences were small and none was statistically significant. The FACIT also had slightly higher correlations with all patient-rated disease specific items from the SHAQ, as well as the adjusted PCS and HAQ-DI, with statistically

significantly higher correlations for items related to RP, gastrointestinal symptoms, breathing and disease severity, as well as the adjusted PCS score and HAQ-DI. Alternatively, the SF-36 Vitality subscale was more closely associated with both mental health measures (adjusted MCS and CES-D), including a statistically significant difference for the adjusted MCS. The correlations with the MCS and PCS are negative because a lower value on the MCS and PCS indicates worse health, whereas higher scores on the fatigue measures represent worse fatigue.

Item and total score discrimination of the SF-36 Vitality subscale and the FACIT

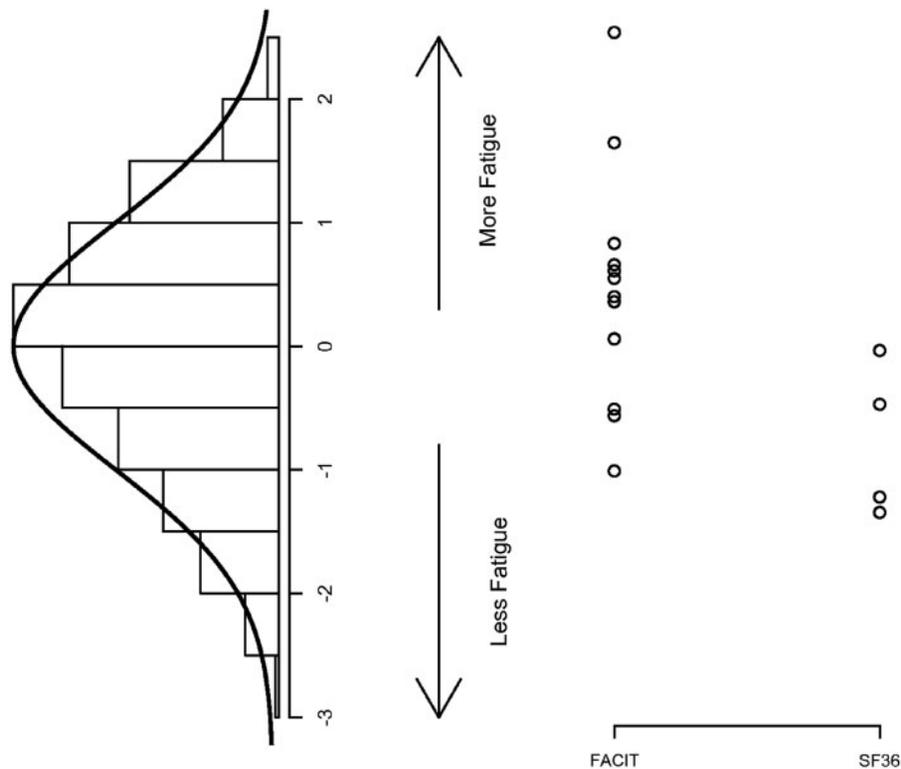
Two assumptions of the PCM model were checked. A confirmatory factor analysis model fitting each instrument separately resulted in an RMSEA index of 0.147 and a BIC value of 309.52, whereas the model fitting both instruments to the same latent trait achieved an RMSEA of 0.119 and BIC of 15.636. Furthermore, fitting a PCM to each instrument separately yielded a BIC value of 14202.04 and fitting a PCM to both instruments together yielded a BIC value of 13807.35. Therefore, there was no evidence that the instruments were measuring different latent constructs, suggesting the unidimensionality. Computation of the measures of manifest monotonicity showed no major violations (>0.3) to the assumption of stochastic ordering.

Fig. 1 shows the degree to which items of the two instruments discriminated between patient fatigue levels. The average severity parameters for the FACIT items covered a larger portion of the fatigue spectrum compared with the SF-36 Vitality subscale, specifically in the range of higher fatigue. Indeed, none of the average severity parameters for the SF-36 Vitality subscale was in the higher levels of fatigue. This suggests that the FACIT items better discriminate amongst patients with moderate to high levels of fatigue than the SF-36 Vitality subscale items. Additional item-level analyses are consistent with this result (see supplementary appendix for details, available at *Rheumatology* Online). Furthermore, when we compared similar plots with Fig. 1 for the diffuse and limited patients separately, the conclusion did not change.

In Fig. 2, the most likely SF-36 Vitality subscale and FACIT scores, based on the most likely item responses at given levels of latent fatigue, are shown. An increasing slope indicates that scores increase with higher levels of fatigue. The horizontal segments, on the other hand, represent ranges of fatigue severity where patients would be expected to have the same score. There are several places where the most likely SF-36 Vitality scores are constant despite substantive changes in fatigue levels (e.g. from –3.3 to –2.2 s.d. below the mean, –0.6 below the mean to 0.1 s.d. above the mean, 0.4 to 1.1 s.d. above the mean). Alternatively, FACIT scores begin to increase for patients –2 s.d. below the mean and increase regularly until fatigue reaches ~2.5 s.d. above the mean. The most likely FACIT score remains constant from –4 s.d. to –3 s.d.

TABLE 2 Associations of SF-36 Vitality subscale and FACIT via Kendall's tau

Covariate	SF-36 Vitality		FACIT		Difference, 95% CI
	r	P	r	P	
Physician global assessments					
Disease severity	0.22	≤0.001	0.26	≤0.001	0.04 (−0.01, 0.09)
Disease activity	0.15	≤0.001	0.18	≤0.001	0.03 (−0.01, 0.09)
Disease damage	0.14	≤0.001	0.16	≤0.001	0.04 (−0.01, 0.09)
Adjusted SF-36 PCS (without Vitality subscale)	−0.37	≤0.001	−0.45	≤0.001	−0.08 (−0.13, −0.03)
Adjusted SF-36 MCS (without Vitality subscale)	−0.38	≤0.001	−0.33	≤0.001	0.05 (0.01, 0.09)
HAQ-DI	0.34	≤0.001	0.40	≤0.001	0.07 (0.02, 0.12)
Patient global assessments					
Pain	0.37	≤0.001	0.40	≤0.001	0.04 (−0.01, 0.09)
RP	0.29	≤0.001	0.34	≤0.001	0.06 (0.00, 0.10)
Finger ulcers	0.10	≤0.014	0.12	≤0.004	0.02 (−0.03, 0.08)
Gastrointestinal problems	0.25	≤0.001	0.30	≤0.001	0.06 (0.01, 0.11)
Breathing	0.37	≤0.001	0.42	≤0.001	0.05 (0.00, 0.10)
Disease severity	0.44	≤0.001	0.51	≤0.001	0.07 (0.03, 0.12)
CES-D	0.49	≤0.001	0.45	≤0.001	−0.04 (−0.08, 0.01)

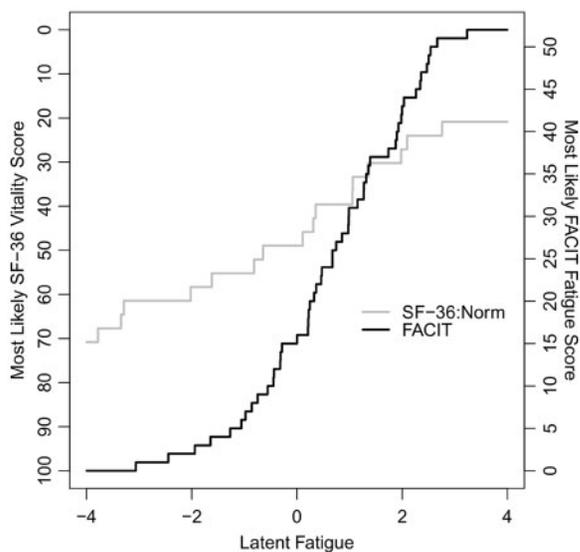
Fig. 1 Severity parameters for the PCM.

Shows the standardized average item severity parameters (across categories) for each item on both instruments in the right panel. The histogram in the left panel shows estimated latent fatigue levels for the 348 patients compared with the standard normal distribution.

below the mean despite substantive changes in fatigue levels. In general, the SF-36 had poor discriminative ability over wide ranges of the fatigue spectrum, and the FACIT was clearly advantageous, particularly in the middle of the

fatigue spectrum (from -2 to 2 s.d. from the mean), changing much more rapidly for changes in the underlying fatigue level. Analysis of the expected scores, rather than most likely scores, yielded similar conclusions (see

Fig. 2 Most likely SF-36 Vitality and FACIT scores as a function of fatigue.



We show the SF-36 norm-based Vitality subscale score (grey line) and the FACIT (black line). This shows the most likely instrument score for a given latent level of fatigue. Note that the left axis is labelled in the original units of the SF-36 Vitality subscale score and the right axis is labelled in the units of the FACIT.

supplementary appendix, Figs A1 and A2, available at *Rheumatology* Online).

Discussion

The main findings of this study were (i) the SF-36 Vitality subscale and the FACIT had similar associations with disease and patient-oriented outcome variables, although the FACIT generally produced somewhat stronger associations; (ii) the individual items of the FACIT provided more discrimination among patients at all but the lowest levels of fatigue, whereas the items of the SF-36 Vitality subscale tended to discriminate well among patients at low to moderate, but not high, levels of fatigue and (iii) the FACIT total score more effectively covered the entire range of fatigue that is encountered in patients with SSc than the SF-36 Vitality subscale.

Fatigue is a common and serious symptom in SSc [9–16] and is the focus of increasing research. Thus, it is imperative that it is measured with a validated instrument with the best possible performance characteristics. The SSc research community has agreed to apply the OMERACT filter to assess the validity of outcomes [60]. The OMERACT filter requires that measures fulfil the following three levels of validation: (i) feasibility; (ii) truth, including face, content, construct and criterion validity and (iii) discrimination, including responsiveness and reliability. The SF-36 Vitality subscale has an advantage over the FACIT in terms of feasibility in that it consists of

4 items versus 13 for the FACIT. The findings of this study suggest that both the SF-36 Vitality subscale and the FACIT are reasonably valid measures of fatigue in SSc. However, the FACIT provides better discrimination across patient fatigue levels in the ranges of the fatigue spectrum commonly observed in patients with SSc and, for that reason, should be the preferred measure of fatigue in SSc. These findings are consistent with results reported by Cella *et al.* [23], who concluded that the FACIT covered a larger portion of the fatigue continuum than the SF-36 Vitality subscale in patients with RA. It is also noteworthy to point out that the FACIT may provide better discrimination across the higher ranges of the fatigue spectrum because while the SF-36 Vitality subscale only asks questions that are general, the FACIT includes items that assess fatigue in everyday situations, such as eating, social activities and the completion of tasks.

There are a number of limitations that should be considered in evaluating the results from this study. The study was based on a convenience sample of patients with SSc. The sample included in this study tended to have a stable pattern of disease, as indicated by the mean duration of 11.1 years since the onset of the first non-RP symptoms. Patients who are not treated by a rheumatologist, who have more severe SSc so as to limit participation in the CSRG Registry or who die early in the course of their disease may be under-sampled in the CSRG Registry. Therefore, our sample may have included an over-representation of healthier patients. On the other hand, our patient sample was drawn from a large number of centres across Canada, and the demographic and disease characteristics of the patient sample in this study are consistent with other outpatient samples reported in the literature [61, 62]. Another limitation of this study is the moderate sample size. Although in SSc research a sample size of 348 would not be considered small, IRT models require many parameters to be estimated. This suggests that there may be some degree of uncertainty in the individual parameter estimates, especially for score categories that did not have a high rate of endorsement. However, this does not alter the main conclusions about the general validity of the SF-36 Vitality subscale and FACIT, as well as the fact that the FACIT covers a larger range of fatigue, as these findings were based on robust patterns in the data. Furthermore, we chose to compare the SF-36 Vitality subscale with the FACIT instead of another measure, such as the MAF because of sample size limitations. Further analyses could be completed to explore the use of other instruments in patients with SSc. A final limitation relates to the nature of the measures that were compared. The SF-36 Vitality subscale is only composed of 4 items, whereas the FACIT has 13 items, such that the data from the FACIT could have been weighted more heavily in estimating the underlying fatigue latent construct in models with both instruments. The wording of the two instruments also differs in temporal period of reference. The SF-36 asks patients to report on symptoms of fatigue over the past 4 weeks, whereas the FACIT asks

for a report over the past 7 days. Because fatigue levels for most patients are not expected to change dramatically due to the general stability, this inconsistency between the wording of the questionnaires should be noted but is not expected to affect the results considerably.

In conclusion, both the SF-36 Vitality subscale and FACIT are reasonably valid measures of fatigue in patients with SSc. The SF-36 Vitality subscale performs well in patients with lower levels of fatigue, but does not discriminate well among patients with moderate to high levels of fatigue. On the contrary, although the FACIT was not able to discriminate well between patients with very low levels of fatigue, it is the preferable instrument for patients with average to high levels of fatigue. Because many patients with SSc experience high levels of fatigue, the FACIT should be the preferred instrument to measure fatigue in this disease.

Rheumatology key messages

- The FACIT had higher correlation with disease characteristics than the SF-36 Vitality subscale in SSc patients.
- The FACIT had better discrimination across part of the fatigue spectrum of interest in patients with SSc.
- The FACIT should be the preferred instrument among patients with SSc.

Acknowledgements

The funding sources had no role in the design of the study, analysis of the data, preparation of the manuscript and decision to submit for publication.

Funding: This study was supported by a grant from the Fonds de la Recherche en Santé du Québec (Principal Investigator, Thombs: #14409) and grants to support the CSRG from the Canadian Institutes of Health Research, the Scleroderma Society of Canada and its provincial chapters, the Scleroderma Society of Ontario, Sclérodémie Québec, the Ontario Arthritis Society, and educational grants from Actelion Pharmaceuticals and Pfizer Inc.

Disclosure statement: The authors have declared no conflicts of interest.

References

- 1 Siebold J. Scleroderma. In: Harris E, Budd R, Firestein G, Genovese M, Sargent J, Ruddy S, eds. *Kelley's textbook of rheumatology*. 7th edn. Philadelphia: Elsevier, 2005.
- 2 Chiffot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. *Semin Arthritis Rheum* 2008;37:223–35.
- 3 Steen VD, Medsger TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 1997;40:1984–91.
- 4 Bernatsky S, Hudson M, Panopalis P *et al*. The cost of systemic sclerosis. *Arthritis Rheum* 2009;61:119–23.
- 5 Mayes MD, Lacey JV Jr, Beebe-Dimmer J *et al*. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum* 2003;48:2246–55.
- 6 Thombs B, Taillefer S, Hudson M, Baron M. Depression in patients with systemic sclerosis: a systematic review of the evidence. *Arthritis Rheum* 2007;57:1089–97.
- 7 Hudson M, Thombs BD, Steele R *et al*. Health-related quality of life in systemic sclerosis: a systematic review. *Arthritis Rheum* 2009;61:1112–20.
- 8 Swain MG. Fatigue in chronic disease. *Clin Sci (Lond)* 2000;99:1–8.
- 9 Bassel M, Hudson M, Taillefer SS *et al*. Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian national survey. *Rheumatology (Oxford)* 2011;50:762–67.
- 10 Van Lankveld WG, Vonk MC, Teunissen H, van den Hoogen FH. Appearance self-esteem in systemic sclerosis: subjective experience of skin deformity and its relationship with physician-assessed skin involvement, disease status and psychological variables. *Rheumatology* 2007;46:872–76.
- 11 Thombs BD, Bassel M, McGuire L *et al*. A systematic comparison of fatigue levels in systemic sclerosis with general population, cancer and rheumatic disease samples. *Rheumatology* 2008;47:1559–63.
- 12 Hudson M, Steele R, Lu Y *et al*. Work disability in systemic sclerosis. *J Rheumatol* 2009;36:2481–6.
- 13 Sandqvist G, Eklund M. Daily occupations-performance, satisfaction and time use, and relations with well-being in women with limited systemic sclerosis. *Disabil Rehabil* 2008;30:27–35.
- 14 Sandqvist G, Scheja A, Eklund M. Working ability in relation to disease severity, everyday occupations and well-being in women with limited systemic sclerosis. *Rheumatology* 2008;47:1708–11.
- 15 Sandusky SB, McGuire L, Smith MT *et al*. Fatigue: an overlooked determinant of physical function in scleroderma. *Rheumatology* 2009;48:165–9.
- 16 Thombs BD, Hudson M, Bassel M *et al*. Sociodemographic, disease, and symptom correlates of fatigue in systemic sclerosis: evidence from a sample of 659 Canadian Scleroderma Research Group Registry patients. *Arthritis Rheum* 2009;61:966–73.
- 17 Ware JE. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;473–83.
- 18 Belza BL, Henke C, Yelin E, Epstein W, Gilliss C. Correlates of fatigue in older adults with rheumatoid arthritis. *Nurs Res* 1993;42:93–9.
- 19 Ibn Yacoub Y, Amine B, Bensabbah R, Hajjaj-Hassouni N. Assessment of fatigue and its relationships with disease-related parameters in patients with systemic sclerosis. *Clin Rheumatol* 2012;31:655–60.
- 20 Sandqvist G, Archenholtz B, Scheja A, Hesselstrand R. The Swedish version of the Multidimensional Assessment of Fatigue (MAF) in systemic sclerosis: reproducibility and

- correlations to other fatigue instruments. *Scand J Rheumatol* 2011;40:493-94.
- 21 Rannou F, Poiraudou S, Berenzne A *et al.* Assessing disability and quality of life in systemic sclerosis: Construct Validities of the Cochin Hand Function Scale, Health Assessment Questionnaire (HAQ), Systemic Sclerosis HAQ, and Medical Outcomes Study 36-Item Short Form Health Survey. *Arthritis Rheum* 2007;57: 94-102.
 - 22 Cella D. Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System. Evanston: Evanston Northwestern Healthcare and Northwestern University. Version 4, 1997.
 - 23 Cella D, Yount S, Sorensen M *et al.* Validation of the functional assessment of chronic illness therapy fatigue scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005;32:811-9.
 - 24 Furst DE, Clements PJ, Steen VD *et al.* The modified Rodnan skin score is an accurate reflection of skin biopsy thickness in systemic sclerosis. *J Rheumatol* 1998;25: 84-8.
 - 25 LeRoy EC, Black C, Fleischmajer R *et al.* Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;15:202-5.
 - 26 Symmons DP. Disease assessment indices: activity, damage and severity. *Baillieres Clin Rheumatol* 1995;9: 267-85.
 - 27 Medsger TA Jr. Natural history of systemic sclerosis and the assessment of disease activity, severity, functional status, and psychologic well-being. *Rheum Dis Clin North Am* 2003;29:255-73.
 - 28 Medsger TA Jr, Silman AJ, Steen VD *et al.* A disease severity scale for systemic sclerosis: development and testing. *J Rheumatol* 1999;26:2159-67.
 - 29 Valentini G, Silman AJ, Veale D. Assessment of disease activity. *Clin Exp Rheumatol* 2003;21(3 Suppl 29): S39-41.
 - 30 Liang MH, Socher SA, Roberts WN, Esdaile JM. Measurement of systemic lupus erythematosus activity in clinical research. *Arthritis Rheum* 1988;31:817-25.
 - 31 Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003;49:156-63.
 - 32 Clements PJ, Seibold J, Furst DE *et al.* High-dose versus low-dose penicillamine in early diffuse systemic sclerosis trial: Lessons Learned. *Semin Arthritis Rheum* 2004;33: 249-63.
 - 33 Hudson M, Impens A, Baron M *et al.* Discordance between patient and physician assessments of disease severity in systemic sclerosis. *J Rheumatol* 2010;37: 2307-12.
 - 34 Fan X, Pope J, Baron M. What is the relationship between disease activity, severity and damage in a large Canadian systemic sclerosis cohort? Results from the Canadian Scleroderma Research Group (CSRG). *Rheumatol Int* 2010;30:1205-10.
 - 35 Ware JE, Kosinski M, Dewey JE. How to score Version Two of the SF-36® Health Survey. Lincoln, RI: QualityMetric Inc., 2000.
 - 36 Dauphinee SW, Gauthier L, Gandek B, Magnan L, Pierre U. Readyng a US measure of health status, the SF-36, for use in Canada. *Clin Invest Med* 1997;20: 224-38.
 - 37 Yellen SB, Cella D, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the functional assessment of cancer therapy (FACT) measurement system. *J Pain Symptom Manag* 1997;13: 63-74.
 - 38 Signorovitch J, Brainsky A, Grotzinger K. Validation of the FACIT-fatigue subscale, selected items from FACT-thrombocytopenia, and the SF-36v2 in patients with chronic immune thrombocytopenia. *Qual Life Res* 2011; 20:1737-44.
 - 39 Hudson M, Steele R, Lu Y *et al.* Clinical correlates of self-reported physical health status in systemic sclerosis. *J Rheumatol* 2009;36:1226-9.
 - 40 Steen VD, Medsger TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 1997;40: 1984-91.
 - 41 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23: 137-45.
 - 42 Van Tubergen A, Debats I, Ryser L *et al.* Use of a numerical rating scale as an answer modality in ankylosing spondylitis-specific questionnaires. *Arthritis Rheum* 2002; 47:242-8.
 - 43 Ferraz MB, Quaresma MR, Aquino LR, *et al.* Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *J Rheumatol* 1990;17: 1022-4.
 - 44 Johnson S, Hawker G, Davis A. The health assessment questionnaire disability index and scleroderma health assessment questionnaire in scleroderma trials: an evaluation of their measurement properties. *Arthritis Care Res* 2005;52:256-62.
 - 45 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385-401.
 - 46 Radloff LS, Locke BZ. Community surveys of psychiatric disorders. In: Weisman NM, Myers JK, Ross CE, eds. *The Community Mental Health Assessment Surveys and the CES-D scale*. New Brunswick: Rutgers University, 1986.
 - 47 Thombs BD, Hudson M, Schieir O, Taillefer SS, Baron M. Reliability and validity of the center for epidemiologic studies depression scale in patients with systemic sclerosis. *Arthritis Rheum* 2008;59:438-43.
 - 48 Efron B, Tibshirani R. *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC, 1993.
 - 49 Marsh H, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: first- and higher order factor models and their invariance across groups. *Psy Bull* 1985;562-82.
 - 50 Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;90:773-95.
 - 51 Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978;6:461-4.

- 52 Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociol Meth Reas* 1992;21:230-58.
- 53 Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982;47:149-74.
- 54 Molenaar IW, Sijtsma K. User's Manual MSP5 for Windows. Groningen: IEC ProGAMMA The Netherlands, 2000.
- 55 Van der Ark L. A. Mokken (version 2.6): an R package for Mokken scale analysis [computer software] 2011. <http://cran.r-project.org/web/packages/mokken/index.html> (15 December 2011, date last accessed).
- 56 R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2010.
- 57 Canty A, Ripley B. Boot: Bootstrap R (S-Plus) Functions. R package version 1.2-43; 2010.
- 58 Davison AC, Hinkley DV. Bootstrap methods and their applications. Cambridge: Cambridge University Press, 1997.
- 59 Rizopoulos D. ltm: An R package for latent variable modeling and item response theory analyses. *J Stat Softw* 2006;17:1-25.
- 60 Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198-9.
- 61 Mayes MD. Scleroderma epidemiology. *Rheum Dis Clin North Am* 2003;29:239-54.
- 62 Singh M, Clements P, Furst D, Maranian P, Khanna D. Work productivity in scleroderma: analysis from the University of California, Los Angeles scleroderma quality of life study. *Arthrit Care Res* 2012;64:176-83.