Support Care Cancer (2011) 19:1441–1450
DOI 10.1007/s00520-010-0971-1

ORIGINAL ARTICLE

Self-reported fatigue: one dimension or more? Lessons from the Functional Assessment of Chronic Illness Therapy—Fatigue (FACIT-F) questionnaire

David Cella · Jin-Shei Lai · Arthur Stone

Received: 12 May 2010 / Accepted: 27 July 2010 / Published online: 13 August 2010

Abstract Across two general population (total n=1,878) and two cancer (total n=3,140) samples, we evaluated the dimensionality of self-reported fatigue as measured by the Functional Assessment of Chronic Illness Therapy—Fatigue (FACIT-F) instrument. After evaluating dimensionality of the FACIT-F, we compared the conceptually distinct fatigue *experience* versus fatigue *impact* scores in each sample. Confirmatory factor analysis of the 13-item scale showed very good fit to a single dimension ("unidimensional") model for each sample (comparative fit index range=0.92–0.97). Using a bifactor model to compare the loading of each item with the general fatigue factor versus the identified sub-domain (experience or impact), we found the item-general loading to be higher than that of the item-sub-domain factor in 52 of 52 comparisons (13 items; four samples). When scored separately, experience and impact scores were correlated highly (range=0.80–0.88), yet their difference relative to one another was significant ($p<$ 0.001). Consistently across samples, experience scores were systematically higher (more endorsement) than impact scores, by a margin of 0.21–0.46 SD units. This suggests that the fatigue experience and the impact of fatigue upon function are reported along a single dimensional continuum, but that *experience* is more likely than *impact upon function* to be endorsed at lower levels of fatigue. Fatigue as an outcome or trial endpoint can be expressed as a single number, and the experience of the symptom is more likely to be endorsed at mild levels of fatigue, presumably before the symptom exerts an adverse impact upon function.

Keywords FACIT-F · Bifactor model · Patient Reported Outcomes Measurement Information System (PROMIS) Cooperative Group

D. Cella (✉) · J.-S. Lai
Northwestern University Feinberg School of Medicine,
Chicago, IL, USA
e-mail: d-cella@northwestern.edu

A. Stone
Stony Brook University,
Stony Brook, NY, USA

Introduction

The Functional Assessment of Cancer Therapy—Fatigue (FACIT-F) is a commonly used and well-validated measure of fatigue in people with cancer and other chronic health conditions [1–6]. Recent reviews of the content of the instrument from US regulatory authorities have suggested that the 13 questions fall into two definable components of fatigue: fatigue experience and fatigue impact (i.e., the impact of fatigue upon activities). Many consider these two components to be distinct dimensions of fatigue. Indeed, investigators of the NIH Roadmap effort known as the Patient Reported Outcomes Measurement Information System (PROMIS) Cooperative Group (www.nihpromis.org) identified fatigue experience and fatigue impact as two *conceptual* dimensions of fatigue, and therefore identified, developed, and otherwise amassed over 100 questions to measure fatigue experience (symptom reporting) and impact of fatigue on activity. The PROMIS investigators entered wave one testing in 2006–2007 with a presumption that experience and impact would evolve into two separately calibrated item banks (which are similar to FACIT-F items, in fact, some PROMIS items are derived from the FACIT-F), each expressing a score on its dimension [7]. Their first wave results, however, indicated that response of people to fatigue experience and fatigue impact questions fell onto the same measurement continuum that they can

collectively be regarded as having sufficient unidimensionality for collapsing into one item bank [8]. Nevertheless, majorities of the PROMIS Wave I samples were recruited from a US general population based pool, and the question remains as to whether the same conclusion could be drawn on disease-specific populations such as cancer population, and whether impact scores would be different from experience scores, if the two were computed separately from FACIT-F data. This paper examines this question in depth, using four available FACIT-F data sets to address the question.

Methods

Data description

FACIT-F data from four archived databases were analyzed, including anemic cancer patients (AC), general cancer patients (GC), an internet sample drawn from the US general population (GP), and a mixed US general population and clinical sample studied in the first wave of PROMIS fatigue item bank testing (PROMIS). AC data consisted of 2,402 mixed-diagnosis cancer patients enrolled in an open-label, non-randomized, community-based clinical trial evaluating the effectiveness, safety, and clinical outcome of a treatment for anemia in cancer patients [1]. The FACIT-F was administered at baseline and multiple follow-up points; only baseline data were analyzed in this study. All patients were receiving some form of chemotherapy at baseline, and all had hemoglobin values equal to or less than 11.0 g/dl (i.e., anemic).

GC data consisted of 738 cancer patients who were recruited from oncology clinics in the Chicago metropolitan for several studies conducted by our research group [2–4]. Broad inclusion criteria were used across these studies, in which sample could have any types of cancer, at any stage, and received any type of treatments.

GP data were collected by Knowledge Networks (Menlo Park, CA, USA), a marketing information and decision support system. Knowledge Networks drew a random sample of 1,075 people, aged 18 and older, from more than 100,000 individuals who were members of an Internet-based survey panel. The panel was a demographically representative sample of the general US adult population. Members of the survey panel responded to one survey per month in exchange for free installation of WebTV Internet service. The FACIT-F (version 4) was one such survey that was presented electronically to the panel members who completed the survey in their homes [5].

The sample from PROMIS (the NIH initiative; http://www.nihpromis.org) was collected by either an internet panel survey research company (Polimetrix now YouGov, Palo Alto, CA, USA), or directly by PROMIS participating clinical sites (University of North Carolina, University of Pittsburgh, Stanford University, and Duke University) for various diseases such as cancer, heart disease, mental illness, and chronic obstructive pulmonary disease. Participants were excluded when (1) average response time was less than one second per item; (2) response time was less than half a second for at least ten consecutive items; or (3) more than 50% missing data were identified. A total of 803 participants completed the FACIT-F items and are included in this report.

Instrument

The FACIT-F is a 13-item questionnaire that assesses self-reported tiredness, weakness, and difficulty conducting usual activities due to fatigue [2]. A five-point intensity type of rating scale (from “not at all” to “very much”) is used. Respondents are instructed to answer questions with respect to their experiences and functioning over the previous 7 days. The FACIT-F is a psychometrically sound instrument and has been widely used to measure fatigue for patients with various chronic illnesses [6] as well as for the US general population [3]. These 13 items were classified into either “experience” (five items) or “impact” (eight items), in order to further examine their relationship to each other. Experience items inquire about perceptions and severity of feeling states, including tiredness, energy level, weakness, fatigue, and listlessness. Impact items inquire as to whether fatigue (or tiredness) has had an impact upon an individual’s daily functioning such as performing usual activities, social activity, sleep, and eating. Item classification was done independently by three investigators with 100% agreement, and these classifications were confirmed in subsequent meetings with external groups including investigators from the pharmaceutical industry and the FDA.

Planned analyses

Analyses were conducted in two stages: (1) evaluating dimensionality of the FACIT-F by using two factor analytic techniques, confirmatory factor analysis (CFA), and bifactor analysis; and (2) investigating the relationship between experience and impact scores.

Evaluating dimensionality of the FACIT-F Given the unique characteristics and large sizes of each sample, each dataset was analyzed separately, and results were then compared. We first evaluated dimensionality of all 13 items that comprise the FACIT-F items by using CFA, specifying a single factor model. If the model fits the data well and the loadings of all of the items are sufficiently large (i.e., >0.3),

the scale can be considered to be unidimensional. To evaluate fit of the CFA model, we applied common criteria for acceptable fit statistics. These include root mean squared error of approximation (RMSEA) *below* 0.10, as well as Tucker–Lewis index (TLI), and comparative fit index (CFI) *above* 0.90 [9, 10].

The bifactor model was used to evaluate the presence of essential unidimensionality of the entire 13 items, thus testing a model driven by clinical observations [4]. This model includes two classes of factors: a general factor (i.e., overall fatigue) (defined by loadings from all of the items in the FACIT-F) and sub-domain factors (i.e., experience and impact) (defined by loadings from pre-specified groups of items related to that sub-domain) [11, 12]. The bifactor model emphasizes factor loadings (i.e., >0.3) [4, 12] to determine whether items are *essentially unidimensional* for measurement applications. If factor loadings for all the items on the general factor are greater than 0.3, and higher on the general factor than they are on the sub-domains, the general factor can then be considered to be measurable, even in the presence of sub-domain factors. Similarly, if the magnitudes of the loadings of all the items on a sub-domain factor are large enough (e.g., >0.3), this would indicate that the sub-domain is measurable, even in the presence of the general factor.

Both CFA and bifactor modeling were conducted using MPlus (version 3) [13] with the implementation of the polychoric correlation matrix and weighted least squares with adjustments for mean and variance estimation, which is appropriate for the evaluation of ordered categorical data.

Relationship between experience and impact We evaluated the degree of association between experience and impact using Pearson correlation coefficients. A scatter plot was also used to further demonstrate this relationship. To examine level differences between experience and impact, paired *t* tests were computed. These analyses were conducted at the database level as well as at individual variable level. With independent *t* tests, we also examined whether participants with different levels of identified clinical variables we believed could be associated with fatigue (e.g., performance status or hemoglobin level) reported similar levels of fatigue experience, fatigue impact, and overall fatigue. Given the nature of this secondary data analysis, not all clinical/demographic information was available for all datasets. The Karnofsky performance rating was used for AC; anemic status, defined as Hgb< 12 g/dl, was used for GC; Eastern Cooperative Oncology Group Performance Status Rating (ECOG PSR) was used for GC; for GP and PROMIS samples, the common general health question was used ("In general, would you say your health is...", 1=excellent to 5=poor). Three additional questions were used in the PROMIS data set: "In general, would you say your quality of life is..." ("QOL"; 1= excellent and 5=poor); "In general, how would you rate your physical health?" ("physical health"; 1=excellent, 5= poor); and "How would you rate your fatigue on average?" ("fatigue"; 1=none and 5=very severe). Analyses were conducted with SAS 9.0 [14].

Results

Sample characteristics

Table 1 shows sample demographic and clinical information by dataset. Across the four datasets, average age ranged from 45.9 (standard deviation=16.6; GP) to 63.3 (standard deviation=12.9; AC). There were more females than males, and most participants were Caucasian. Most participants rated themselves as having good performance status (AC, Karnofsky≥70; GC dataset, ECOG=0 or 1), good or better health (available for PROMIS and GP), good or better quality of life (available for PROMIS), and good or better physical health (available for PROMIS). In the AC group, average hemoglobin value was 9.29 g/dl (standard deviation=1.03), and 23.3% had lung cancer, followed by breast (16.0%), gynecologic (12.6%), non-Hodgkin lymphoma (10.0%), and gastro-intestinal cancer (8.7%). For GC, the average hemoglobin value was 12.02 g/dl, and 34.7% had breast cancer, followed by gastro-intestinal (14.4%), lung (8.4%), non-Hodgkin's lymphoma (7.7%), and gynecologic cancer (7.5%).

Table 2 reports internal consistency reliability (alpha) as well as mean, standard deviation, and selected percentile scores (1, 25, 50, 75, and 99) for each sample on the FACIT-F experience, impact, and total scores. It also presents, at the bottom of the table, internal consistency coefficients for each FACIT-F subscale on each sample, if each item were to be removed from the scale. The internal consistency of each FACIT scale in each sample is so high (range 0.88–0.96) that removal of any one item rarely has a significant effect on improving (or worsening) internal consistency by its removal. This is an indirect support of the notion that each scale is essentially reporting on the same concept.

Factor analysis results

CFA and bifactor analysis results are shown in Table 3. For the three CFA analyses (testing one factor model for general fatigue, fatigue impact, and fatigue experience), CFI and TLI were greater than 0.9 across four databases. However, RMSEA values were all greater than 0.10 except impact in PROMIS (RMSEA=0.08) and experi-

Table 1 Sample demographic and clinical information

	Anemic cancer (AC: n=2,402)	PROMIS (n=803)	General population (GP: n=1,075)	General cancer (GC: n=738)
Mean age	63.3 (SD=12.9)	51.8 (SD=17.8)	45.9 (SD=16.58)	58.7 (SD=13.55)
Hemoglobin value (g/dl)	9.29 (SD=1.03)	NA	NA	12.02 (SD=1.72)
Gender				
Female	59.37%	54.99%	51.67%	64.30%
Male	40.63%	45.01%	48.33%	35.84%
Race (endorsed all apply)				
Caucasian	78.25%	80.90%	75.90%	86.72%
African American	11.68%	10.11%	10.23%	6.91%
Hispanic origin	5.29%	10.90%	10.04%	2.71%
Performance status				
Karnofsky 90 or 100 or ECOG=0	32.56%			37.62%
Karnofsky 70 or 80 or ECOG=1	48.84%			41.59%
Karnofsky<70 or ECOG=2–4	18.59%			20.79%
Health in general				
Excellent		11.72%	19.36%	
Very good		36.91%	36.54%	
Good		34.66%	30.97%	
Fair		14.09%	10.67%	
Poor		2.62%	2.46%	
Quality of life in general				
Excellent		18%		
Very good		42%		
Good		29%		
Fair		10%		
Poor		2%		
Physical health				
Excellent		11%		
Very good		34%		
Good		35%		
Fair		18%		
Poor		3%		
Fatigue on average				
None		16.94%		
Mild		41.97%		
Moderate		34.99%		
Severe		5.73%		
Very severe		0.37%		
Cancer site				
Breast	16.03%			34.71%
Gastro-intestinal	8.70%			14.40%
Lung	23.27%			8.37%
Non-Hodgkin's lymphoma	9.95%			7.68%
Gynecologic	12.61%			7.54%
Prostate	NA			4.12%
Leukemia	3.75%			3.29%
Hodgkin's lymphoma	5.50%			2.47%

PROMIS Patient Reported Outcomes Measurement Information System, *SD* Standard deviation, *ECOG* Eastern Cooperative Oncology Group

Table 2 FACIT-F scores, distributions, and internal consistency across four samples

		Anemic cancer			General cancer			General population			PROMIS		
		Experience	Impact	Total	Experience	Impact	Total	Experience	Impact	Total	Experience	Impact	Total
Alpha mean (SD)		0.920	0.890	0.940	0.920	0.910	0.950	0.900	0.880	0.930	0.930	0.930	0.960
		7.86 (5.26)	15.99 (7.89)	23.85 (12.62)	13.26 (4.87)	23.62 (6.89)	36.88 (11.36)	13.97 (4.78)	26.11 (6.14)	40.07 (10.36)	13.54 (4.60)	25.18 (6.81)	38.73 (11.08)
Percentile[a]	99%	19	32	50	20	32	52	20	32	52	20	32	52
	75%	12	22	34	17	29	46	17	31	48	17	31	48
	50%	7	16	23	14	25	39	15	28	43	15	27	42
	25%	4	10	14	10	19	30	11	23	35	11	21	31
	1%	0	2	2	0	5	6	0	6	9	1	6	8
Alpha if item deleted		(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)	(Alpha)
Experience items													
I feel listless ("washed out")		0.892		0.932	0.896		0.944	0.865		0.921	0.907		0.953
I feel tired		0.890		0.932	0.897		0.944	0.865		0.921	0.906		0.954
I have energy		0.936		0.937	0.926		0.947	0.913		0.926	0.925		0.955
I feel weak all over		0.899		0.934	0.907		0.944	0.875		0.922	0.919		0.954
I feel fatigued		0.897		0.934	0.897		0.944	0.866		0.922	0.903		0.953
Impact items													
I am too tired to eat			0.892	0.943		0.907	0.950		0.874	0.930		0.928	0.959
I need help doing my usual activities			0.876	0.939		0.902	0.949		0.871	0.930		0.922	0.957
I am frustrated by being too tired to do the things I want to do			0.868	0.935		0.892	0.945		0.849	0.921		0.912	0.953
I have to limit my social activity because I am tired			0.862	0.933		0.891	0.944		0.853	0.923		0.913	0.953
I have trouble *starting* things because I am tired			0.863	0.932		0.892	0.944		0.853	0.921		0.911	0.952
I have trouble *finishing* things because I am tired			0.862	0.933		0.892	0.944		0.850	0.920		0.914	0.953
I am able to do my usual activities			0.885	0.940		0.906	0.949		0.874	0.929		0.920	0.955
I need to sleep during the day			0.890	0.942		0.905	0.949		0.877	0.930		0.927	0.958

[a] Experience, impact, and total fatigue scores are standardized to a range of 0–4 for ease of comparison. To compute raw scores comparable with FACIT manual, multiply mean or percentile scores by the number of items (experience×5; impact×8; total fatigue×13)

Alpha Cronbach's coefficient alpha after the item was removed, *FACIT-F* Functional Assessment of Chronic Illness Therapy—Fatigue, *SD* Standard deviation

Table 3 Factor analysis summary

		Anemic cancer	General cancer	General population	PROMIS
CFA general fatigue	CFI	0.929	0.926	0.923	0.973
	TLI	0.982	0.985	0.977	0.994
	RMSEA	0.187	0.162	0.176	0.119
CFA impact	CFI	0.965	0.962	0.966	0.993
	TLI	0.983	0.985	0.980	0.997
	RMSEA	0.173	0.151	0.178	0.080
CFA experience	CFI	0.996	0.992	0.996	0.997
	TLI	0.997	0.992	0.997	0.997
	RMSEA	0.128	0.162	0.096	0.113
Bifactor	CFI	0.975	0.955	0.980	0.990
	TLI	0.994	0.99	0.994	0.998
	RMSEA	0.108	0.134	0.088	0.075

CFA confirmatory factor analysis, *CFI* comparative fit index, *TLI* Tucker–Lewis index, *RMSEA* root mean square error of approximation

ence in GP (RMSEA=0.096). All item factor loadings were greater than 0.5. Testing the bifactor model, CFI and TLI were excellent, exceeding 0.95 across all datasets. RMSEA was less than 0.10 for two general population data, but not for the two cancer patient datasets.

Individual item-to-factor loadings for each analysis are presented in Table 4. All items had higher loadings on the general factor (i.e., fatigue), ranging from 0.56 ("I need to sleep during the day" in AC) to 0.95 ("I have trouble starting things because I am tired" in AC and PROMIS), than on the sub-domain (i.e., experience and impact), ranging from −0.31 ("I have trouble starting things because I am tired" in GP) to 0.54 ("I feel fatigued" in GP).

We further examined the relationship between experience and impact by examining their correlations with one another and comparing their performance in known-groups validity tests. Correlations and paired *t* test results are shown in Table 5. Across the analyses for each of the datasets, experience-to-impact correlation coefficients ranged from 0.80 (GP) to 0.88 (PROMIS). Within each dataset, at the level of available demographic variables, correlation coefficients ranged from 0.79 (male in GP) to 0.89 (age<65 in PROMIS). These consistently high correlations support the conclusion that the experience of fatigue is so highly correlated with its impact that separate reporting may be redundant. However, an interesting and

Table 4 Comparisons of the item loadings from bifactor analysis (overall fatigue versus sub-domain) across all databases

Item		Anemic cancer		General cancer		General population		PROMIS	
		General fatigue	Sub-domain	General fatigue	Sub-domain	General fatigue	Sub-domain	General fatigue	Sub-domain
I feel listless ("washed out")	Experience	0.84	0.42	0.86	0.51	0.82	0.41	0.87	0.27
I feel tired		0.87	0.37	0.85	0.22	0.81	0.41	0.86	0.37
I have energy		0.73	0.09	0.79	0.06	0.71	0.09	0.86	0.02
I feel weak all over		0.80	0.42	0.87	0.23	0.79	0.42	0.85	0.25
I feel fatigued		0.82	0.38	0.85	0.20	0.75	0.54	0.89	0.37
I am too tired to eat	Impact	0.62	0.00	0.78	−0.08	0.74	0.16	0.80	−0.15
I need help doing my usual activities		0.69	0.27	0.76	−0.23	0.65	0.38	0.82	−0.16
I am frustrated by being too tired to do the things I want to do		0.81	0.28	0.88	−0.03	0.87	0.26	0.94	0.00
I have to limit my social activity because I am tired		0.86	0.38	0.88	−0.05	0.82	0.31	0.91	−0.10
I have trouble starting things because I am tired		0.95	−0.24	0.90	0.39	0.93	−0.31	0.95	0.27
I have trouble finishing things because I am tired		0.92	−0.11	0.87	0.22	0.93	−0.21	0.90	0.17
I am able to do my usual activities		0.64	0.18	0.75	−0.28	0.65	0.19	0.84	−0.15
I need to sleep during the day		0.56	0.06	0.69	−0.11	0.58	0.15	0.71	−0.02

All items had higher loading on the general fatigue than their local factors across all dataset, supporting essential unidimensionality of the FACIT-F

Table 5 Comparisons between fatigue experience (EXP) and impact (IMP) scores

		Anemic cancer (n=2,402)	General cancer (n=738)	General population (n=1,075)	PROMIS (n=803)
All sample together	Correlation paired t test (EXP–IMP)	0.83	0.86	0.80	0.88
		$t = -34.4 (p < 0.001)$	$t = -16.42 (p < 0.001)$	$t = -26.9 (p < 0.001)$	$t = -27.6 (p < 0.001)$
Age ≥65	Correlation paired t test (EXP–IMP)	0.83 (n=1,139)	0.85 (n=257)	0.82 (n=164)	0.87 (n=193)
		$t = -25.27 (p < 0.001)$	$t = -10.76 (p < 0.001)$	$t = -9.91 (p < 0.001)$	$t = -11.58 (p < 0.001)$
<65	Correlation paired t test (EXP–IMP)	0.85 (n=1,130)	0.88 (n=473)	0.80 (n=909)	0.89 (n=547)
		$t = -23.72 (p < .001)$	$t = -12.5 (p < .001)$	$t = -25.0 (p < .001)$	$t = -25.4 (p < .001)$
Gender Male	Correlation paired t test (EXP–IMP)	0.84 (n=890)	0.84 (n=260)	0.79 (n=516)	0.88 (n=331)
		$t = -22.79 (p < 0.001)$	$t = -10.28 (p < 0.001)$	$t = -17.77 (p < 0.001)$	$t = -18.09 (p < 0.001)$
Female	Correlation paired t test (EXP–IMP)	0.83 (n=1,386)	0.87 (n=470)	0.80 (n=544)	0.88 (n=408)
		$t = -25.87 (p < 0.001)$	$t = -12.85 (p < 0.001)$	$t = -20.23 (p < 0.001)$	$t = -20.82 (p < 0.001)$

potentially important observation revealed in these analyses was a systematic shift in the level of fatigue severity when comparing separate scores for experience and impact. Specifically, impact scores, although highly related to experience scores, were systematically higher (better) than experience scores. On the standardized 0–4 scale, the differences between average experience scores and average impact scores were as follows: anemic cancer, 0.43 (SD= 0.59); general cancer, 0.30 (SD=0.50); general population, 0.47 (SD=0.57); and PROMIS, 0.44 (SD=0.43). The effect sizes of these differences between experience and impact range from 0.61 (general cancer) to 1.01 (PROMIS), suggesting medium to large effects. In addition, paired t test comparisons of impact versus experience were significantly different across all comparisons, $p<0.001$ (see Table 5). These results indicate that while fatigue experience and impact are highly correlated, and can be measured as a unidimensional concept, the nature of that strong relationship is such that fatigue experience raw scores will be systematically lower (worse) than fatigue impact raw scores.

To further illustrate this relationship between experience and impact, we plotted the experience scores against impact scores. As shown in Figs. 1, 2, 3, and 4, fatigue experience and fatigue impact were highly correlated with slopes of regression lines ranging from 0.89 (anemic cancer) to 1.00 (GP), however, with different degrees of fatigue severity intercepts ranging from −0.21 for anemic cancer to −0.46 for GP.

Mean comparisons for participants in different performance rating and/or anemic status are shown in Table 6. Responses to impact, experience, and overall fatigue were examined separately in order to examine whether these three types of scoring discriminated levels of fatigue in a similar manner. All comparisons were statistically significant at p=0.01, indicating that experience alone, impact alone, and overall fatigue (i.e., all 13 FACIT-F items) could statistically discriminate participants with different performance levels and/or anemic status.

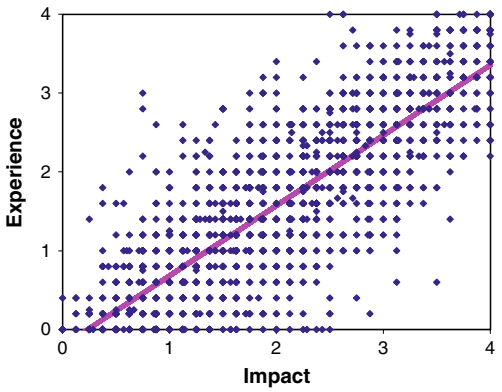

Fig. 1 Fatigue experience versus fatigue impact—anemic cancer (AC: n=2,402). Experience = (−0.21) + 0.89 × Impact

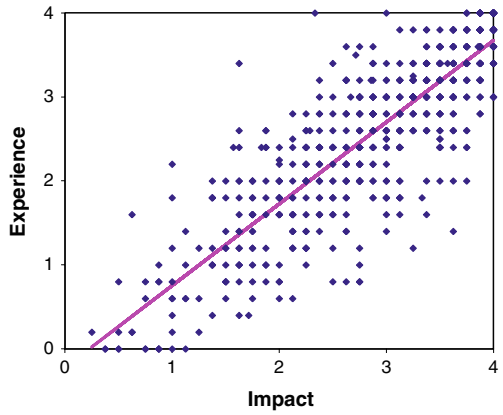

Fig. 2 Fatigue experience versus fatigue impact—general cancer patients (GC: n=738). Experience = $(-0.22) + 0.97 \times$ Impact

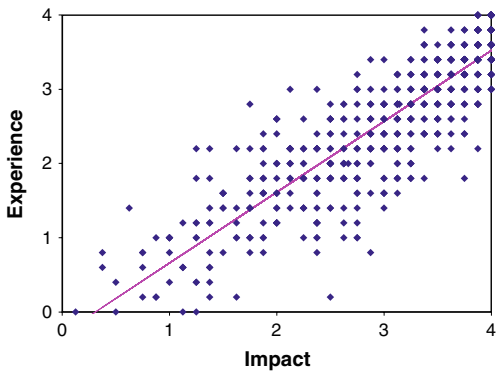

Fig. 4 Fatigue experience versus fatigue impact—general population (PROMIS; n=803). Experience = $-0.29 + 0.95 \times$ Impact

Discussion

We evaluated the dimensionality of self-reported fatigue as measured by the FACIT-F instrument. Confirmatory factor analysis of the 13-item scale showed very good fit to a unidimensional model for each sample. Using a bifactor model to compare the loading of each item with the general fatigue factor versus the identified subdomain (experience or impact), we found the item-general loading to be higher than that of the item-subdomain factor in 52 of 52 comparisons. Separate experience and impact scores were very highly correlated, yet their difference relative to one another was significant (p<0.001). Consistently across samples, experience scores were systematically higher (more endorsement) than impact scores, by a margin of 0.21–0.46 SD units. This strongly suggests that the fatigue experience and the impact of fatigue upon function are reported along a single dimensional, but that experience is likely to be endorsed at lower levels than impact upon function. Practically speaking, this implies that fatigue as an outcome or trial endpoint can be expressed as a single number, and the experience of the symptom is more likely to be endorsed at mild levels of fatigue, presumably before the symptom exerts an adverse impact upon function.

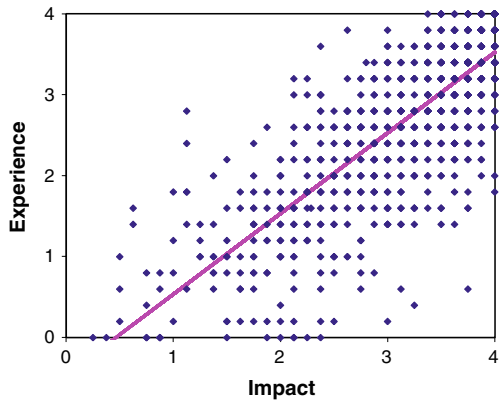

Fig. 3 Fatigue experience versus fatigue impact—general population (GP; n=1,075). Experience = $-0.46 + 1.00 \times$ Impact

Despite its composition of five experience and eight impact questions, the total FACIT-F score is a reasonable and parsimonious endpoint choice for clinical research. Results were consistent across four distinct data sets, using confirmatory factor analytic techniques (CFA and bifactor model analysis). Despite the conceptual distinction between fatigue experience and fatigue impact, results confirmed that fatigue experience and impact are aligned on the same dimension. By implication, reporting of a single fatigue score is reasonable and even preferred from a measurement efficiency standpoint. When calibrated using item response theory modeling, which takes the high covariation between experience and impact into account, all 13 items can be treated as reflecting a common underlying concept. However, in circumstances where the researcher would prefer to report experience or impact scores separately, this can be supported by the bifactor model and CFA results for each sub-domain. Nevertheless, scores on each component (experience and impact) will be expressed on the same single measurement continuum.

Based upon regulatory review experience, some investigators from the pharmaceutical industry have expressed interest in reporting fatigue by separating experience from impact. Given this interest, we chose the bifactor model analysis approach for this study. Bifactor analysis is applicable when the following three conditions apply: (1) There is a general factor that is hypothesized to account for the commonality of the items; (2) there are multiple domains specific factors (they are named local factors, or sub-domains in this study), each of which is hypothesized to account for the unique influence of the local factors over and above the general factor; and (3) researchers may be interested in the local factors as well as the general factor that is of focal interest. Following condition 2, the relationship between general and local factors is assumed

Table 6 Group differences on selected variables between "experience," "impact," and "overall fatigue" scores

Database	Variable	Mean comparison (group 1–group 2)a, [b, c, d, e, f, g]		
		Experience	Impact	Overall fatigue
AC	Karnosky[a]	−0.48*	−0.62*	−0.57*
GC	Anemic status[b]	0.34*	0.30*	0.32*
	ECOG PSR[c]	−1.20*	−1.20*	−1.21*
GP	Health[d]	−0.97*	−0.84*	−1.05*
PROMIS	Health[d]	−1.13*	−1.10*	−1.11*
	QOL[e]	−1.24*	−1.23*	−1.24*
	Physical function[f]	−1.07*	−1.00*	−1.03*
	Fatigue[g]	1.18*	1.03*	1.09*

*$p<0.001$

AC anemic cancer patient, *GC* general cancer patients, *GP* general population dataset, PROMIS

[a] Karnosky: Range from 10 to 100, in which 10=moribund and 100=normal. In the analysis, participants who endorsed 10–60 (group=1) were compared with those who endorsed 70–100 (group=2)

[b] Anemic status: Anemia when hemoglobin value <12 g/dl (group=2); non-anemic when Hgb≥12 g/dl (group=1)

[c] *ECOG PSR* Eastern Cooperative Oncology Group Performance Status rating, in which 0=normal activity, 1=normal activity with some symptom, 2=confined to bed <50% of waking time, 3=confined to bed >50% of waking time, and 4=unable to get out of bed. In the analysis, participants who endorsed ratings of 0 or 1 (group=2) were compared with those endorsed ratings 2–4 (group=1)

[d] Health: In general, would you say your health is 5=excellent, 4=very good, 3=good, 2=fair, 1=poor. In the analysis, participants endorsed ratings of 1–2 (group=1) were compared with those who endorsed 3–5 (group=2)

[e] QOL: In general, would you say your quality of life is: 5=excellent, 4=very good, 3=good, 2=fair, 1=poor. In the analysis, participants endorsed ratings of 1–2 (group=1) were compared with those who endorsed 3–5 (group=2)

[f] Physical function: In general, how would you rate your physical health? 5=excellent, 4=very good, 3=good, 2=fair, 1=poor. In the analysis, participants endorsed ratings of 1–2 (group=1) were compared with those who endorsed 3–5 (group=2)

[g] Fatigue: How would you rate your fatigue on average? 1=none, 2=mild, 3=moderate, 4=severe, 5=very severe. In the analysis, participants endorsed ratings of 1 or 2 (group=2) were compared with those who endorsed 3–5 (group=1)

to be orthogonal [15]. Results of the bifactor analysis confirmed our hypothesis that from a measurement's perspective, FACIT-F can be reported as a single score. However, it can also be reported as two separate scores, experience and impact, if clinicians prefer to do so as acceptable fit indices of the hypothesized model with these two sub-domains were found. Two additional considerations could equally support selection of one or the other of the fatigue components (experience and impact) or both. One is that the CFA model fit was slightly better for the individual components compared with the general model. The second is that experience scores, while highly correlated to impact scores, appeared to be systematically lower (worse) than impact scores, across the samples. Results indicated that FACIT-F is unidimensional and it is valid to report a single score. Meanwhile, according to bifactor analysis, it would be valid if clinicians decide to report experience and impact scores separately.

Some limitations regarding this research help point to future compelling research directions. For example, further research is needed to determine if experience and impact scores vary in different ways across different clinical groups, or if these two components change in different ways over time with treatment. If indeed either of these differences is shown in future research, a case could be made for recording scores separately. Until such time, however, single score reporting seems most efficient for clinical research purposes. In addition, the 7-day recall period may play a role in the common reporting of experience impact. Averaging experience and impact over a week—or whatever people do in the reporting process—may yield scores that spuriously covary to a degree at this level. Research that examines the relationship between experience and impact based upon momentary or 24-h recall would help inform this issue.

Compared with non-cancer samples, cancer patient data showed higher RMSEA than those of general population. RMSEA is typically used to provide an intuitive summary of discrepancy between the population and the model subspace by rescaling them to a comparable metric. Higher RMSEA values among cancer patients might indicate that fatigue is experienced slightly more diversely in its heterogeneous nature (in terms of fatigue severity from none to severe) than fatigue experienced by general population (i.e., more homogeneous). Regardless, acceptable CFI and TLI suggest the essential unidimensionality

with and without taking the local factors of experience and impact into account.

These results indicated fatigue experience and fatigue impact lie on the same trait, but at different severity levels, with experience "shifted" toward more severity. This is a common observation in measurement, in which items are calibrated on different locations on the same measurement continuum depending on the degrees of severities that item measures, in this case, level of fatigue. This type of observations has facilitated the gradual popularity of the item bank in health-related quality of life and its applications such as computerized adaptive testing (CAT) [16]. From a clinical perspective, a reasonable interpretation is that fatigue is typically experienced before it has an impact upon function. It therefore stands to reason that one would be more likely to report fatigue experience before reporting an impact upon functioning. It would be worthwhile to evaluate whether this interpretation is applicable in examining change in fatigue over time. When faced with situations or treatments that impact fatigue, it may be that changes in experience precede changes in impact, and in such cases, examining both experience and impact would provide the most complete understanding of patients' fatigue.

It also appears likely that for people with mild fatigue, experience questions may be more responsive to difference or change than impact (since few with mild fatigue will report significant impact). Alternatively, for those in the moderate to severe range of fatigue, impact questions may be more responsive to difference or change than experience questions. Given this, the best approach to measuring fatigue across its continuum of severity is by use of a combination of experience and impact questions, pooling the responses together into one total score, and reporting either one single fatigue score or two separate scores depending on clinical need. In settings such as PROMIS where there is an item bank of fatigue experience and impact questions residing in the same unidimensional bank, a previously mentioned CAT approach would be likely to select experience items more often for people with low levels of fatigue and impact items for people with higher levels of fatigue.

In the setting of static fatigue assessment such as with the FACIT-F, it appears that combining of both types of questions is advisable. However, if one were to decide to split experience and fatigue items and produce separate scores, this practice can be justified by these results. Such would be case when the construct of experience and impact are of theoretical or applied importance; for instance, treatments intended to help patients better manage their fatigue might be shown to mainly affect fatigue impact and little on fatigue experience. However, it is likely that the fatigue impact measure would be slightly more vulnerable to ceiling effects, and perhaps the fatigue experience more vulnerable to floor effects (though this is less likely), using this approach compared with the full FACIT-F.

Conflict of interest statement None.

References

1. Demetri GD, Kris M, Wade J, Degos L, Cella D (1998) Quality-of-life benefit in chemotherapy patients treated with epoetin alfa is independent of disease response or tumor type: results from a prospective community oncology study. Procrit Study Group. J Clin Oncol 16:3412–3425
2. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E (1997) Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage 13:63–74
3. Cella D, Lai JS, Chang CH, Peterman A, Slavin M (2002) Fatigue in cancer patients compared with fatigue in the general United States population. Cancer 94:528–538
4. Lai JS, Crane PK, Cella D (2006) Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. Qual Life Res 15:1179–1190
5. Cella D, Zagari MJ, Vandoros C, Gagnon DD, Hurtz HJ, Nortier JW (2003) Epoetin alfa treatment results in clinically significant improvements in quality of life in anemic cancer patients when referenced to the general population. J Clin Oncol 21:366–373
6. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J (2005) Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol 32:811–819
7. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M (2007) The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH roadmap cooperative group during its first two years. Med Care 45:S3–S11
8. Lai JS, Cella D, Choi S, Teresi JA, Hays RD, Stone AA (2008) Developing a fatigue item bank for the Patient-Reported Outcomes Measurement Information System (PROMIS FIB version 1). Presented at the Meeting of the Survey Methods in Multicultural, Multinational, and Multiregional Contexts (3MC), Berlin, Germany.
9. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 6:1–55
10. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D (2007) Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 45:S22–S31
11. Gibbons R, Hedeker D (1992) Full-information item bi-factor analysis. Psychometrika 57:423–436
12. McDonald RP (1999) Test theory: a unified treatment. Lawrence Earlbaum, Mahwah
13. Muthen LK, Muthen BO (2004) MPlus: statistical analysis with latent variables. Muthen & Muthen, Los Angeles
14. S. A. S. Institute Inc. (2003) SAS/STAT, Version 9.1.
15. Chen FF, West SG, Sousa KH (2006) A comparison of bifactor and second-order models of quality of life. Multivariate Behav Res 41:189–225
16. Cella D, Gershon R, Lai JS, Choi S (2007) The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res 16:133–141