Scandinavian Journal of Gastroenterology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/igas20

The multidimensional fatigue inventory (MFI-20): psychometrical testing in a Norwegian sample of inflammatory bowel disease (IBD) patients

Lars-Petter Jelsness-Jørgensen, Bjørn Moum, Tore Grimstad, Jørgen Jahnsen, Øistein Hovde, Svein Oskar Frigstad & Tomm Bernklev

To cite this article: Lars-Petter Jelsness-Jørgensen, Bjørn Moum, Tore Grimstad, Jørgen Jahnsen, Øistein Hovde, Svein Oskar Frigstad & Tomm Bernklev (2022): The multidimensional fatigue inventory (MFI-20): psychometrical testing in a Norwegian sample of inflammatory bowel disease (IBD) patients, Scandinavian Journal of Gastroenterology, DOI: 10.1080/00365521.2022.2029939

To link to this article: https://doi.org/10.1080/00365521.2022.2029939

Published online: 25 Jan 2022.

SCANDINAVIAN JOURNAL OF GASTROENTEROLOGY
https://doi.org/10.1080/00365521.2022.2029939

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

The multidimensional fatigue inventory (MFI-20): psychometrical testing in a Norwegian sample of inflammatory bowel disease (IBD) patients

Lars-Petter Jelsness-Jørgensen[a,b], Bjørn Moum[c,d], Tore Grimstad[e,f], Jørgen Jahnsen[d,g], Øistein Hovde[d,h], Svein Oskar Frigstad[d,i] and Tomm Bernklev[d,j]

[a]Faculty of Health, Welfare and Organisation, Østfold University College, Halden, Norway; [b]Department of Gastroenterology, Østfold Hospital Trust, Grålum, Norway; [c]Department of Gastroenterology, Oslo University Hospital, Nydalen, Norway; [d]Institute of Clinical Medicine, University of Oslo, Blindern, Norway; [e]Department of Gastroenterology, Stavanger University Hospital, Stavanger, Norway; [f]Department of Clinical Science, University of Bergen, Bergen, Norway; [g]Department of Gastroenterology, Akershus University Hospital, Lillestrøm, Norway; [h]Department of Gastroenterology, Innlandet Hospital Trust, Gjøvik, Norway; [i]Department of Medicine, VestreViken Baerum Hospital, Gjettum, Norway; [j]Vestfold Hospital Trust, Tønsberg, Norway

ABSTRACT

Objectives: To evaluate the psychometric properties of the Norwegian version of the multidimensional fatigue inventory (MFI-20) in patients with inflammatory bowel disease.

Methods: Participants were recruited from nine hospitals in the southeastern and western parts of Norway. Clinical and sociodemographic data were collected, and participants completed the MFI-20, as well as the Fatigue Questionnaire (FQ). In addition to a confirmatory factor analysis, validity, reliability, test-retest and responsiveness were evaluated.

Results: In total, 410 patients were included. The Norwegian MFI-20 had an acceptable model fit when compared to the original five-dimensional structure. A positive correlation was observed between the dimensions of MFI-20 and the FQ. MFI-20 scores increased according to subjective disease activity, but no differences were observed when using a calprotectin cut-off $<$ or $>=$250 µg/g mg/kg. All MFI-20 dimensions except 'reduced motivation' in both ulcerative colitis (UC) and Crohn's disease (CD) patients had alpha Cronbach alpha values $\geq$70, and test-retest reliability revealed good to excellent values. Merely one dimension (Reduced activity) in UC patients reporting improvement did not reach the threshold for acceptable responsiveness according to Guyatt statistics.

Conclusions: The Norwegian version of MFI-20 is valid, reliable and responsive. The instrument can safely be used in studies using fatigue as an endpoint.

ARTICLE HISTORY
Received 30 October 2021
Revised 2 January 2022
Accepted 11 January 2022

KEYWORDS
Inflammatory bowel disease; fatigue; psychometrical testing; multidimensional fatigue inventory

Introduction

Inflammatory bowel disease (IBD) is a collective term encompassing ulcerative colitis (UC) and Crohn's disease (CD). Both diseases are typically characterized by chronic, recurrent inflammation of the gastrointestinal (GI) tract [1–3]. Predominant IBD symptoms include diarrhea with or without blood and mucus, as well as abdominal cramping, pain and fatigue [3–5]. In UC the inflammation is located to a varying extent of the colonic mucosa, while the entire GI wall and any part of GI-tract, from mouth to anus, may be affected in CD [2,3,6].

In recent years there has been an increased awareness of fatigue as a central symptom among IBD patients [5,7–10]. Studies have shown that fatigue is more prevalent in IBD patients than in controls [5,11,12], and that it negatively affects health-related quality of life (HRQoL) and other psychosocial issues [13–15]. While no universal definition of fatigue exists, it may be referred to as an overwhelming sense of physical and mental exhaustion that is not alleviated with adequate rest and is distinct from depression. Currently only a limited number of tools to measure fatigue has been adequately psychometrically tested in IBD [7]. To our knowledge, merely two questionnaires, the Facit-F and the disease specific IBD-F, have been fully tested [16,17]. Jelsness-Jørgensen et al. [5] reported the validity of the Fatigue Questionnaire (FQ) in a study of chronic fatigue, and later on the test-retest reliability of the same questionnaire in a subsequent letter [18]. The FQ's sensitivity to change remains, however, unknown.

As pointed out in the review by Czuber-Dochan et al. [7], the Multidimensional Fatigue Inventory (MFI-20) is the questionnaire most commonly used to measure fatigue in IBD. The Norwegian version of the MFI-20 has undergone forward and backward translation, as well as linguistic validation in a prior study in fibromyalgia patients [19]. It has, however, not been psychometrically tested, neither in a Norwegian – nor an IBD population. The aim of the current study was consequently to evaluate the psychometric properties of the Norwegian MFI-20 in IBD patients.

CONTACT Lars-Petter Jelsness-Jørgensen lars.p.jelsness-jorgensen@hiof.no
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

Materials and methods

Patients were consecutively recruited as part of a large multicenter study, consisting of nine different hospitals in the south-eastern and western part of Norway. All patients had to be ≥18 years of age, have a verified diagnosis of IBD, based on endoscopic, laboratory and histological findings, and all had to be able to read and write Norwegian and willing to give informed consent. Patients were not included if they had cognitive impairment, or were judged by the investigators to be unlikely to comply with the study procedures. At each of the including centers a senior gastroenterologist was in charge of the study.

Socio-demographic and clinical data

Socio-demographic variables were collected directly from the patients and included age, gender, civil status, educational level, work status and smoking habits.

Data regarding clinical status, symptoms, and current use of medications were obtained through laboratory tests, calprotectin levels in stools (FeCal-test), disease activity indices (SCCAI/SCDAI), clinical investigation and medical records. In addition, patients classified their IBD symptoms present during the last 14 days, providing four possible scores: no symptoms, mild symptoms (do not interfere with everyday activities), moderate symptoms (do interfere with everyday activities, may result in sick leave), and severe symptoms (unable to carry out everyday activities, on sick leave, or hospitalized). Disease phenotype was described according to the Montreal Classification.

Questionnaires

The multidimensional fatigue inventory (MFI-20)

The Multidimensional Fatigue Inventory (MFI-20) is a 20-item, self-report instrument designed to measure five dimensions of fatigue (general fatigue, physical fatigue, reduced activity, reduced motivation, and mental fatigue) [20]. Each question is scored from one to five and each dimension consists of five questions. The dimensional score consequently ranges from 4 to 20 (a higher score indicates more fatigue). Permission to use the MFI-20 in this study was kindly given by professor Ellen M.A. Smets at the Academic Medical Center (AMC), University of Amsterdam.

The fatigue questionnaire (FQ)

The FQ was developed by Chalder et al. [21]. Symptoms related to Chronic Fatigue Syndrome (CFS) were not included in the scale as the primary purpose was to develop a scale that measured solely fatigue. The FQ has been translated into Norwegian and validated [22]. Studies have demonstrated that the questionnaire has stable psychometric properties across populations. The FQ consists of 11 items, divided into two main dimensions: Physical Fatigue (PF; 7 items) and Mental Fatigue (MF; 4 items). Responses lay along a continuum with four options used (0 = better than usual, 1 = no more than usual, 2 = worse than usual, and 3 = much worse than usual). Higher scores imply higher levels of fatigue. Combining the scores of PF and MF produce Total Fatigue (TF), with a maximum scale score of 33. In addition, the FQ contains two questions involving the duration and extent of fatigue symptoms. The scale scores of the FQ are also scored on a dichotomized scale (0 = better than usual and no more than usual, 1 = worse than usual, and much worse than usual). Based on the results of the original validation study, the Norwegian validation study, and general consensus, chronic fatigue (CF) was defined as dichotomized scores >=4 and a duration >6 months [21,22].

Analysis

To assess the characteristics of the sample we used descriptive analysis, frequencies and the χ^2 test. Confirmatory factor analysis, including standardized root mean square residual – SRMR, root mean square error of approximation – RMSEA, and comparative Fit Index – CFI, were used to assess how well the original five-dimensional model fitted the data [23–25]. *Convergent validity* was calculated using binary correlation analysis (Spearman's rho) of the MFI-20 and the FQ. It was hypothesised that elevated fatigue (increased MFI-20 scores) would correlate positively with all FQ dimensions, but that the strongest correlations would be identified between the dimensions general and physical fatigue in MFI-20 and the physical fatigue dimension of the FQ. A similar relationship was hypothesized to be identified between the mental fatigue dimension in MFI-20 and the mental fatigue dimension in the FQ. *Known group validity* was tested through one-way analysis of variance (ANOVA), by comparing mean MFI-20 scores in patients reporting no, mild, moderate or serious IBD symptoms. A fecal calprotectin cut-off of < or >= 250 mg/kg were also used to investigate the differences in MFI-20 scores between those categorized as having active disease or not [26]. *Floor and ceiling* effects were investigated by calculating the percentage of patients scoring either the lowest or highest possible score in individual items as well as in dimensional scores. If the number of lowest or highest possible scores on the MFI-20 exceeded 15%, this was, according to recommendations [27], regarded as indicative of floor or ceiling effects. *Internal consistency reliability* was tested with Cronbach's alpha. *Test-retest reliability* was measured using the intraclass correlation coefficient (ICC, two-way mixed, single measure) between baseline and follow-up (four to six weeks apart). At follow-up, patients self-reported whether or not their IBD condition was unchanged, improved or deteriorated. Based on this item, ICC values were calculated among those patients reporting to be in a stable condition. *Responsiveness* was calculated by comparing the MFI-20 scores on baseline to those after 4–6 weeks in patients that reported either worsening or improvement in IBD symptoms. Both a Guyatt's statistic and Cohens'd effect size were used. Guyatt's statictic was performed by dividing the mean change in individuals reporting either improvement or deterioration of symptoms with the standard deviation of the change score in those unchanged [28]. If the Guyatt statistic was greater than 1.00 (or −1.00) for those

patients that reported either deterioration or improvement, it was considered as highly responsive to change, whereas a value greater than 0.20 (or −0.20) was considered acceptable [28]. Cohen's d effect size was calculated by comparing the mean difference between groups, divided by the pooled standard deviation. Operational definitions of 0.2, 0.5, and 0.8 were categorized as small, medium, and large, respectively. Missing data were treated as recommended in the literature; if data in half or less than half of the items within a scale were missing, they were replaced by the mean value of the respondent's completed items in the same scale [29]. All tests were 2-sided, with a 5% significance level and performed by the use of Predictive Analytics Software, PASW, version 27.0 (SPSS Inc. 233 S. Wacker Drive, Chicago, Illinois – United States) and IBM AMOS, version 27.

Ethical considerations

Participation in the study was based on verbal information from the responsible gastroenterologist followed by written informed consent from the patient and performed in accordance with the principles of the revised Helsinki Declaration. Approval was obtained from the Regional Ethics Committee (reference number: 2012/845/REK Sør-Øst A).

Results

A total of 452 patients were eligible and invited to participate. Four hundred and fourteen patients (91.6%) gave written informed consent. Four of these patients were excluded since the number of missing data exceeded 50%, leaving the number included for analyses at 410. Of these, 230 were diagnosed with CD and 180 with UC. Baseline characteristics of the included patients are presented in Table 1. There were no significant differences in gender or age between patients either declining participation, being excluded due to missing values or those included in analyses. Data on the diagnosis of those declining participation was however, not available.

After inviting all 410 patients from baseline to complete the MFI-20 a second time, a total of 243 responded, corresponding to 59% of the original sample (CD 130/230, UC 113/180). None of those 243 patients responding at the retest had missing values on the MFI-20. In CD however, one responder had not indicated whether or not his/her condition was unchanged, improved or deteriorated. Hence a total of 110 CD patients reported that their condition was unchanged compared to baseline, while 14 reported symptom improvements and five deterioration. The comparable numbers in UC were 86 unchanged, 20 improved and eight deteriorated. In neither UC nor CD, floor or ceiling effects exceed 15%. There was as significant positive correlation between the patients' subjective classification of IBD symptoms during the last 14 days, the SCCAI total score (.63 $p < .001$) and SCDAI total score (.61 $p > .001$), respectively.

Validity and responsiveness

Confirmatory factor analysis revealed an acceptable model fit, with an SRMR, RMSEA and CFI value of 0.06, 0.08 and 0.88, respectively (Figure 1). There was a positive correlation between all dimensions of the MFI-20 and the FQ, of which the strongest correlations aligned with our predefined hypothesis (Table 2). Known group validation revealed elevated MFI dimensional scores as the patients' subjective IBD symptoms increased (Table 3). No differences in fatigue scores were observed according to calprotectin cut-off.

In CD patients reporting either an improved or deteriorated condition, all Cohen's d effect sizes were higher than 0.20. In UC patients reporting improvement – figures were lower than 0.20 in the dimensions 'physical fatigue' and 'reduced activity', while in UC patients reporting deterioration – figures were lower than 0.20 in the dimensions 'reduced activity' and 'mental fatigue'. Except for 'reduced

Table 1. Sociodemographic and clinical characteristics according to diagnosis.

	UC ($n = 180$)	CD ($n = 230$)	p Value
Age mean (SD)	40.8 (12.6)	40.7 (13.0)	ns
Age (range)	18–76	18–77	ns
Gender			
Female	87	114	
Male	93	116	ns
Time since diagnosis (years)	8.8 (8.2)	13.6 (10.5)	<.001
SCCAI total score	3.4 (3.1)		
SCDAI total score		4.7 (3.8)	
UC extent¶			
E1-proctitis	20 (11.1%)		
E2-left-sided colitis	58 (32.2%)		
E3-extensive colitis	102 (56.7%)		
CD localization¶			
L1-terminal ileum (+L4)		75 (32.6%) (8 (25.0%))	
L2-colon (+L4)		47 (20.4%) (6 (18.8%))	
L3-ileocolon (+L4)		76 (33.0%) (18 (56.3))	
L4-upper GI		32 (13.9%)	
CD behavior¶			
B1-nonpenetrating/nonstricturing (+p)		117 (50.9%) (23 (19.7%))	
B2-penetrating (+p)		30 (13.0%) (13 (43.3%))	
B3-stricturing (+p)		83 (36.1%) (15 (18.1%))	

UC: ulcerative colitis; CD: Crohn's disease; SD: standard deviation; SCCAI: Simple Clinical Colitis Activity Index; SCDAI: Simplified Crohn's Disease Activity Index; ¶: Montreal classification; ns: nonsignificant, +p: perianal disease. Figures are in mean and standard deviation if not otherwise noted.

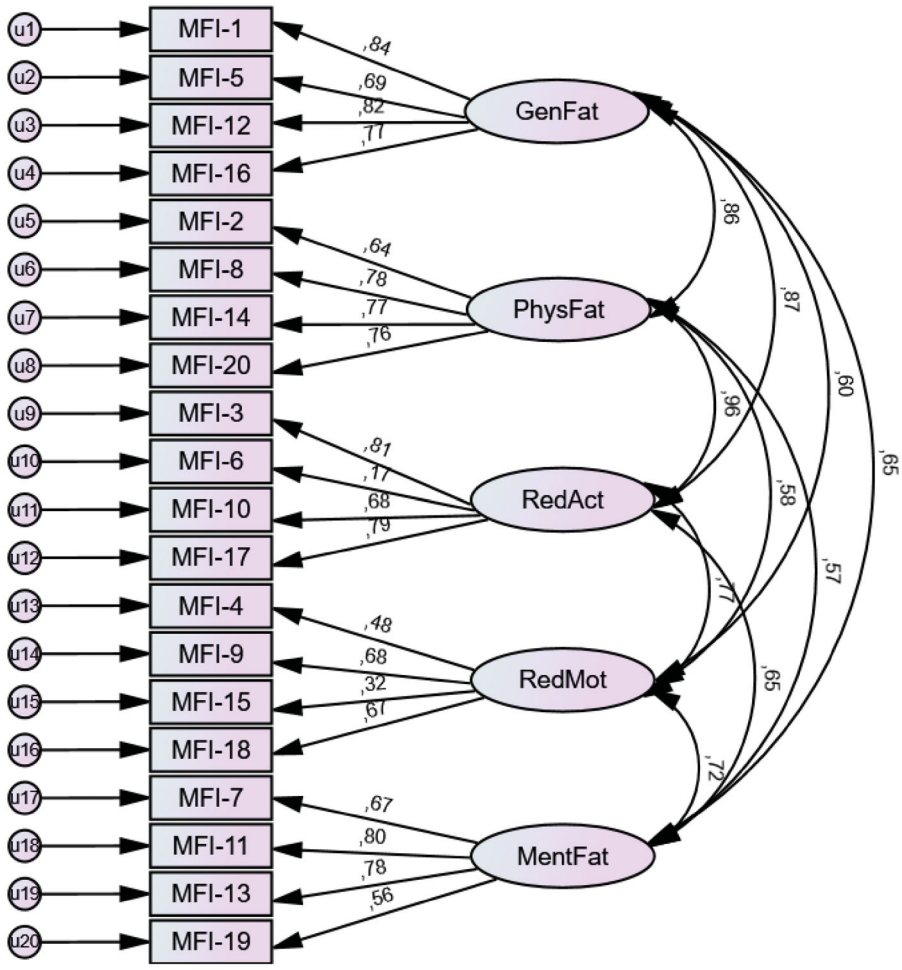

Figure 1. Confirmatory Factor Analysis of the Norwegian version of the MFI-20. Ellipses represent the dimensions from the original MFI-20, rectangles represent the items of the MFI-20 and circles represent measurement error. Two-headed arrows represent correlation coefficients and one-headed arrows factor loadings.

activity' in UC patients reporting improvement, the Guyatt statistic revealed values ≥0.20 (Table 4).

Reliability

Calculation of internal consistency revealed that Cronbach alphas ranged from 0.64–0.84 in UC and 0.65–0.87 in CD. All MFI-20 dimensions except 'reduced motivation' in both UC and CD had alpha values ≥70. Intraclass correlation in patients reporting an unchanged condition between baseline and retest is presented in Table 4.

Discussion

This study set out to evaluate the psychometric properties of the MFI-20 in a Norwegian population of IBD patients. Our findings indicate that the Norwegian version of this widely used fatigue instrument has an acceptable model fit, is valid, reliable and responsive to change.

The use of patient reported outcome measures (PROMs) in clinical research has increased substantially over the last decades. To draw valid conclusions based on such outcomes, it is of outmost importance that measurement tools such as the MFI-20 is able to demonstrate sufficiently robust psychometric properties. Indeed, the International Society for Quality of Life Research (ISOQOL) has recommended minimum standards for the selection of PROMs used in patient outcome research [30]. The aim is of course to enhance the rigor, quality and usefulness of this research. Although the standards are primarily aimed at the original development and testing of questionnaires, they also address the evaluation of translated versions. However, while the 'equivalence of measurement properties for translated versions' was not included in the recommended minimum standards, data on these properties was regarded as desirable.

As pointed out by Hooper et al. [23] there is a plethora of indices to evaluate model fit. They conclude that it is pointless to report everyone, but that at the same time one should not be tempted to choose those who show the best model fit. The selected fit-indices in the current study was consequently based on these recommendations and the fact that the SRMR, RMSEA and CFI values reflect different aspects of model fit. Furthermore, cut-off values for the different indices vary, and the subsequent guidance have also changed over time [23–25]. When we look at the recommendations for the different indices and results in our study, the conclusion is that the original 5-dimension MFI-20 model has an acceptable but not optimal fit in relation to our data [23].

Optimally, tools that are assumed to measure the same phenomenon should be correlated [31]. As anticipated

Table 2. Bivariate (Spearman) correlation between the dimensions of the MFI-20 and the FQ.

	Ulcerative Colitis ($n = 180$)			Crohn's disease ($n = 228^{\pm}$)		
	FQ-PF	FQ-MF	FQ-TF	FQ-PF	FQ-MF	FQ-TF
GF	.71	.50	.72	.70	.42	.68
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
PF	.66	.43	.66	.59	.38	.58
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
RA	.58	.43	.58	.48	.37	.49
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
RM	.39	.34	.42	.36	.32	.37
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
MF	.48	.51	.53	.46	.54	.54
	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$

MFI-20: Multidimensional Fatigue Inventory; FQ: Fatigue Questionnaire; $^{\pm}$: missing data ($n = 2$); FQ-PF, MF and TF: Physical, mental and total fatigue scores of the Fatigue Questionnaire; GF, PF, RA, RM, MF: MFI-20 dimensions general fatigue, physical fatigue, reduced activity, reduced motivation and mental fatigue.

increased MFI-20 scores correlated positively with all FQ dimensions, while the strongest associations were identified between those sub dimensions that were hypothesized to be most closely linked. Furthermore, given the lack of a gold-standard comparison, evaluating the questionnaire among respondents expected to score differently, e.g., in those with different disease activity, is one way of ensuring validity. Our results showed that the fatigue scores increased in line with the patients' subjective indication of disease severity. No differences were however found when calprotectin was used as an indicator of disease severity. In our view, the latter is probably not related to a debatable validity, but rather to the fact that the association between fatigue and objective markers of disease activity has been reported as low [12].

Table 3. MFI-20 scores according to self-reported disease activity.

	Ulcerative Colitis ($n = 180$)				Crohn's disease ($n = 230$)			
	No$n = 31$	Mild$n = 65$	Moderate$n = 61$	Severe$n = 23$	No$n = 38$	Mild$n = 79$	Moderate$n = 83$	Severe$n = 30$
GF	10.2 (3.3)	12.1 (4.1)	14.5 (3.5)	16.3 (1.9)	11.8 (4.4)	12.3 (4.3)	14.4 (4.0)	16.0 (3.5)
PF	9.3 (3.9)	11.1 (4.0)	13.5 (3.4)	16.5 (2.8)	10.3 (3.9)	11.5 (4.2)	12.9 (3.9)	15.4 (3.8)
RA	7.9 (2.1)	9.2 (3.4)	10.4 (3.2)	12.4 (3.3)	8.7 (3.2)	9.8 (3.4)	10.4 (3.7)	12.0 (3.4)
RM	7.6 (2.5)	8.6 (3.4)	9.0 (2.7)	10.2 (4.1)	7.7 (2.8)	8.2 (2.8)	8.9 (3.6)	10.2 (3.7)
MF	9.0 (3.2)	10.2 (3.7)	11.1 (3.5)	13.0 (5.3)	8.9 (3.3)	9.4 (3.2)	11.3 (3.8)	12.7 (3.4)

MFI-20: Multidimensional Fatigue Inventory-20; GF: General fatigue; PF: Physical fatigue; RA: Reduced activity; RM: Reduced motivation; MF: Mental fatigue. All figures presented as means with standard deviations.

Table 4. Test-retest reliability, sensitivity and responsiveness of the MFI-20.

	Unchanged condition											
	Ulcerative Colitis ($n = 86$)						Crohn's disease ($n = 110$)					
	V0	V1	ICC	*p*	*d*	G	V0	V1	ICC	*p*	*d*	G
GF	12.7 (4.1)	13.0 (4.4)	0.90	<.001	0.07	−0.16	13.3 (4.4)	13.4 (4.3)	0.88	<.001	0.02	−0.05
PF	11.5 (4.5)	11.7 (4.6)	0.91	<.001	0.04	−0.10	12.3 (4.3)	12.2 (4.4)	0.88	<.001	0.02	0.05
RA	9.3 (3.4)	10.2 (3.7)	0.79	<.001	0.25	−0.40	10.2 (3.6)	10.8 (3.9)	0.83	<.001	0.18	−0.25
RM	8.2 (3.1)	8.6 (3.3)	0.78	<.001	0.09	−0.17	8.8 (3.3)	8.7 (3.3)	0.82	<.001	0.03	0.04
MF	10.3 (4.0)	10.7 (4.2)	0.92	<.001	0.07	−0.25	10.3 (3.5)	10.4 (3.8)	0.88	<.001	0.03	0.08
	Improved condition											
	Ulcerative Colitis ($n = 19$)						Crohn's disease ($n = 14$)					
	V0	V1	ICC	*p*	*d*	G	V0	V1	ICC	*p*	*d*	G
GF	14.0 (5.0)	12.8 (5.3)	0.79	<.001	0.31	0.62	15.2 (3.8)	13.7 (4.2)	0.73	<.01	0.37	0.71
PF	13.0 (4.7)	12.6 (5.0)	0.89	<.001	0.11	0.21	14.2 (4.2)	12.4 (4.2)	0.73	<.001	0.43	0.81
RA	11.2 (2.7)	10.8 (3.7)	0.63	<.01	0.12	0.17	10.7 (3.8)	8.9 (3.0)	0.84	<.001	0.52	0.81
RM	10.3 (3.3)	8.6 (3.5)	0.64	<.01	0.50	0.81	8.9 (3.2)	7.8 (3.2)	0.82	<.001	0.34	0.55
MF	10.2 (4.2)	9.4 (4.0)	0.88	<.001	0.20	0.47	11.8 (4.0)	10.6 (3.9)	0.92	<.001	0.30	0.67
	Deteriorated condition											
	Ulcerative Colitis ($n = 8$)						Crohn's disease ($n = 5$)					
	V0	V1	ICC	*p*	*d*	G	V0	V1	ICC	*p*	*d*	G
GF	13.9 (2.2)	16.0 (1.3)	0.54	.04	0.53	−1.10	13.6 (3.3)	16.2 (3.6)	0.07	.34	0.64	−1.23
PF	14.0 (4.5)	16.5 (3.5)	0.64	.02	0.17	−1.30	13.4 (3.4)	14.4 (4.3)	0.53	.01	0.23	−0.45
RA	9.8 (3.1)	13.1 (2.4)	0.32	.17	0.69	−1.43	11.9 (2.8)	15.6 (3.5)	−0.11	.36	0.91	−1.68
RM	8,1 (3.2)	10.4 (2,6)	0.59	.03	0.49	−1.10	8.6 (4.0)	9.6 (4.2)	−0.14	.29	0.36	−0.50
MF	10.8 (5.2)	13.1 (5.1)	0.78	<.01	0.15	−1.35	9.6 (5.2)	14.0 (4.3)	0.26	.64	0.88	−2.44

MFI-20: Multidimensional Fatigue Inventory-20; GF, PF, RA, RM, MF: MFI-20 dimensions general fatigue, physical fatigue, reduced activity, reduced motivation and mental fatigue; V0/V1: Mean MFI-20 values at baseline and follow-up; ICC: Intraclass correlation coefficient; *p*: statistical significance; *d*: Cohen's d effect size; G: Guyatt statistics.

As highlighted in the minimum standards from ISOQOL, a PROM should have evidence of responsiveness [30]. This include evidence of changes in scores which are consistent with changes in the target population. In other words, if the condition of the IBD patient changes, the MFI-20 scores should follow. A PROMs responsiveness is of course paramount when used in e.g., longitudinal studies. In our study, the number who reported change was low, which may be related to the relatively short test-retest period. Four to six weeks is probably too short a time interval to detect major changes in disease activity. Nevertheless, with the exception of the MFI-20 dimension 'reduced activity' in UC patients who experienced improvement, all Guyatt statistics values were acceptable [28].

A subject of debate has been the number needed to include in test-retest analysis, ranging from 50 and up [32,33]. A total of 196 patients in the current study met these requirements, which we consider robust. Within the selected timeframe of 4 to 6 weeks, the MFI-20 displayed good to excellent values, regardless of diagnosis.

The study is of course not without limitations. Responsiveness could have been measured once more to confirm the observations after 4–6 weeks. In addition, subjective reporting of disease activity is an obvious subject of possible bias. Optimally, we could therefore have used an objective measure of intestinal inflammation to assess changes over time.

In summary, we conclude that the Norwegian version of MFI-20 is valid, reliable and responsive. Consequently, the instrument can safely be used in studies using fatigue as an endpoint.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Østfold Hospital Trust, Østfold University Hospital and an unrestricted grant by Tillotts Pharma.

ORCID

Lars-Petter Jelsness-Jørgensen http://orcid.org/0000-0002-5465-1576
Jørgen Jahnsen http://orcid.org/0000-0002-0419-2914
Svein Oskar Frigstad http://orcid.org/0000-0001-7841-608X

References

[1] Moum B, Ekbom A. Epidemiology of inflammatory bowel disease–methodological considerations. Digest Liver Dis. 2002;34(5):364–369.
[2] Ponder A, Long MD. A clinical review of recent findings in the epidemiology of inflammatory bowel disease. Clin Epidemiol. 2013;5:237–247.
[3] Loftus EV. Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology. 2004;126(6):1504–1517.
[4] Moum B, Vatn MH, Ekbom A, et al. Incidence of crohn's disease in four counties in southeastern Norway, 1990-93. A prospective population-based study. The inflammatory bowel South-Eastern Norway (IBSEN) Study group of gastroenterologists. Scand J Gastroenterol. 1996;31(4):355–361.
[5] Jelsness-Jorgensen LP, Bernklev T, Henriksen M, et al. Chronic fatigue is more prevalent in patients with inflammatory bowel disease than in healthy controls. Inflam Bowel Dis. 2011;17(7):1564–1572.
[6] Satsangi J, Silverberg MS, Vermeire S, et al. The montreal classification of inflammatory bowel disease: controversies, consensus, and implications. Gut. 2006;55(6):749–753.
[7] Czuber-Dochan W, Ream E, Norton C. Review article: description and management of fatigue in inflammatory bowel disease. Aliment Pharmacol Ther. 2013;37(5):505–516.
[8] Farrell D, Artom M, Czuber-Dochan W, et al. Interventions for fatigue in inflammatory bowel disease. Cochrane Database Syst Rev. 2020;4(4):CD012005.
[9] Borren NZ, van der Woude CJ, Ananthakrishnan AN. Fatigue in IBD: epidemiology, pathophysiology and management. Nat Rev Gastroenterol Hepatol. 2019;16(4):247–259.
[10] McGing JJ, Radford SJ, Francis ST, et al. Review article: the aetiology of fatigue in inflammatory bowel disease and potential therapeutic management strategies. Aliment Pharmacol Ther. 2021;54(4):368–387.
[11] Bager P, Befrits R, Wikman O, et al. Fatigue in out-patients with inflammatory bowel disease is common and multifactorial. Aliment Pharmacol Therapeut. 2012;35(1):133–141.
[12] Grimstad T, Norheim KB, Isaksen K, et al. Fatigue in newly diagnosed inflammatory bowel Disease. J Crohns Colitis. 2015;9(9):725–730.
[13] Jelsness-Jorgensen LP, Bernklev T, Henriksen M, et al. Chronic fatigue is associated with impaired health-related quality of life in inflammatory bowel disease. Aliment Pharmacol Ther. 2011;33(1):106–114.
[14] Jelsness-Jorgensen LP, Bernklev T, Henriksen M, et al. Chronic fatigue is associated with increased disease-related worries and concerns in inflammatory bowel disease. World J Gastroenterol. 2012;18(5):445–452.
[15] Opheim R, Fagermoen MS, Bernklev T, et al. Fatigue interference with daily living among patients with inflammatory bowel disease. Qual Life Res. 2014;23(2):707–717.
[16] Tinsley A, Macklin EA, Korzenik JR, et al. Validation of the functional assessment of chronic illness therapy-fatigue (FACIT-F) in patients with inflammatory bowel disease. Aliment Pharmacol Therapeut. 2011;34(11-12):1328–1336.
[17] Czuber-Dochan W, Norton C, Bassett P, et al. Development and psychometric testing of inflammatory bowel disease fatigue (IBD-F) patient self-assessment scale. J Crohns Colitis. 2014;8(11):1398–1406.
[18] Jelsness-Jorgensen LP. The fatigue questionnaire has a good test-retest profile in IBD. Aliment Pharmacol Ther. 2012;35(5):621–622.
[19] Le Gal M, Mainguy Y, Le Lay K, et al. Linguistic validation of six patient-reported outcomes instruments into 12 languages for patients with fibromyalgia. Joint Bone Spine. 2010;77(2):165–170.
[20] Smets EM, Garssen B, Bonke B, et al. The multidimensional fatigue inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995;39(3):315–325.
[21] Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. J Psychosom Res. 1993;37(2):147–153.
[22] Loge JH, Ekeberg O, Kaasa S. Fatigue in the general norwegian population: normative data and associations. J Psychosom Res. 1998;45(1):53–65.
[23] Hooper D, Coughlan J, Mullen MR. Structural equation modelling: Guidelines for determining model fit. Electronic J Business Res Methods. 2008;6(1):53–60.
[24] Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equat Model. 1999;6(1):1–55.

[25] Steiger JH. Understanding the limitations of global fit assessment in structural equation modeling. Personal Individ Differences. 2007;42(5):893–898.
[26] Bressler B, Panaccione R, Fedorak RN, et al. Clinicians' guide to the use of fecal calprotectin to identify and monitor disease activity in inflammatory bowel disease. Can J Gastroenterol Hepatol. 2015;29(7):369–372.
[27] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293–307.
[28] Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171–178.
[29] Ware JE, Kosinski M. SF-36 physical & mental health summary scales: a manual for users of version 1: QualityMetric; 2001.
[30] Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8):1889–1905.
[31] Chin C-L, Yao G. Convergent validity. In: Michalos AC, editor. Encyclopedia of quality of life and Well-Being research. Dordrecht: Springer Netherlands; 2014. p. 1275–1276.
[32] Stucky B, Pereira C, de Vet HCW, et al. Measurement in medicine: a practical guide. Qual Life Res. 2012;21(2):371–373. : Dordrecht
[33] Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. Qual Life Res. 2014; 23(6):1713–1720.