



In search of distinct MS-related fatigue subtypes: results from a multi-cohort analysis in 1.403 MS patients

Gesa E. A. Pust^{1,2} · Jana Pöttgen¹ · Jennifer Randerath^{2,3} · Stephanie Lau¹ · Christoph Heesen^{1,4} · Stefan M. Gold^{1,5,6} · Iris-Katharina Penner^{7,8}

Received: 28 January 2019 / Revised: 2 April 2019 / Accepted: 2 April 2019 / Published online: 19 April 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Fatigue is among the most disabling symptoms in patients with multiple sclerosis (PwMS). The common distinction between cognitive and motor fatigue is typically incorporated in self-rating instruments, such as the Chalder Fatigue Questionnaire (CFQ), the Fatigue Scale for Motor and Cognitive Functions (FSMC) or the Modified Fatigue Impact Scale (MFIS). The present study investigated the factor structure of the CFQ, the FSMC and the MFIS utilizing exploratory (EFA) and confirmatory factor analysis (CFA) as well as exploratory structural equation modeling (ESEM). Data of 1.403 PwMS were analyzed, utilizing four samples. The first sample ($N=605$) was assessed online and split into two stratified halves to perform EFA, CFA, and ESEM on the CFQ and FSMC. The second sample ($N=293$) was another online sample. It served to calculate CFA and ESEM on the CFQ and FSMC. The third sample was gathered in a clinical setting ($N=196$) and analyzed by applying CFA and ESEM to the FSMC. The fourth sample ($N=309$) was assessed in a clinical setting and allowed to run a CFA and ESEM on the MFIS. Proposed factor structures of all questionnaires were largely confirmed in EFA. However, none of the calculated CFAs and ESEMs could verify the proposed factor structures of the three measures, even with oblique rotation techniques. The findings might have implications for future research into the pathophysiological basis of MS-related fatigue and could affect the suitability of such measures as outcomes for treatment trials, presumably targeting specific sub-components of fatigue.

Keywords Fatigue · Assessment · Validation · Multiple sclerosis

Introduction

Fatigue is one of the most common symptoms or comorbidities in many neurological disorders, including stroke, neurodegenerative diseases, and inflammatory disorders of

the CNS [1]. Fatigue is particularly common in multiple sclerosis (MS), where it affects over 70% of patients and it is often perceived as the most disabling symptom [2]. Importantly, fatigue is a major predictor of unemployment in MS [3, 4] and significantly impairs quality of life [5].

✉ Iris-Katharina Penner
Iris-Katharina.Penner@uni-duesseldorf.de

¹ Institut für Neuroimmunologie und Multiple Sklerose (INIMS), Zentrum für Molekulare Neurobiologie Hamburg (ZMNH), Universitätsklinikum Hamburg Eppendorf, Hamburg, Germany

² Department of Psychology, University of Konstanz, Constance, Germany

³ Lurija Institute for Rehabilitation and Health Sciences at the University of Konstanz, Schmieder Foundation for Sciences and Research, Allensbach, Germany

⁴ Klinik und Poliklinik für Neurologie, Universitätsklinikum Hamburg-Eppendorf (UKE), Hamburg, Germany

⁵ Charité Universitätsmedizin Berlin, Klinik für Psychiatrie und Psychotherapie, Campus Benjamin Franklin, Berlin, Germany

⁶ Charité Universitätsmedizin Berlin, Medizinische Klinik m.S. Psychosomatik, Campus Benjamin Franklin, Berlin, Germany

⁷ Department of Neurology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

⁸ COGITO Center for Applied Neurocognition and Neuropsychological Research, Düsseldorf, Germany

Despite the high clinical importance of fatigue in MS, its pathophysiology remains poorly understood [1]. Research to unravel its causes and develop effective treatments remains challenging. Attempts to objectively quantify fatigue using experimental tests or biomarkers have yielded inconclusive results, so that self-report questionnaires remain the clinical gold standard [6]. Due to the large inter-individual heterogeneity in MS pathogenesis and disease manifestations throughout the CNS, subtypes of MS-fatigue with distinct neurobiological signatures have been suggested [7]. This approach could help to dissect the complex architecture of MS-fatigue but would require measures to differentiate subtypes or components of this syndrome.

The first scales that were developed have typically defined fatigue as a uni-dimensional construct (Fatigue Severity Scale; [8]; Fatigue Impact Scale; [9]). However, tools such as the Fatigue Scale for Motor and Cognitive Functions (FSMC; [10]), the Chalder Fatigue Questionnaire (CFQ; [11]) and the Modified Fatigue Impact Scale (MFIS; [12]) differentiate motor and cognitive fatigue, largely based on clinical observations and the related utility of this distinction. The latter scale, MFIS, has well-documented psychometric shortcomings [13, 14]. The FSMC has been validated on a large sample of patients with MS (PwMS) and healthy controls and distinguishes cognitive from motor fatigue [10]. The validity of the CFQ has also been demonstrated in the general population [15]. It assesses fatigue severity and a distinction between mental and physical fatigue. Both the FSMC and the CFQ have good psychometric properties in PwMS (reliability, validity, and sensitivity) [10, 16].

However, the psychometric validity of distinct fatigue subconstructs remains statistically doubtful. Studies of Martin et al. [15] and Chilcot et al. [16] revealed that the proposed factor structure of the CFQ could not be replicated by means of confirmatory factor analyses (CFA) and that items needed to be excluded to obtain an acceptable model-fit. Likewise, a recent study by Picariello et al. [17] demonstrated that the factor structure as well as the item number of the CFQ had to be adapted to meet fit criteria. Despite the thorough and scientifically sound development of the CFQ and FSMC, the underlying factor structures of the two scales as well as the MFIS have, to the best of our knowledge, not yet been confirmed by CFA or exploratory structural equation modeling (ESEM).

The goal of the present study was to evaluate the factorial validity of subconstructs of MS-fatigue utilizing exploratory factor analyses (EFA), CFA and ESEM. To this end, we leveraged several independent, large data sets to determine whether validly discernible subtypes of fatigue exist on the psychometric level in MS.

Methods

Participants

Data of 1,403 PwMS, derived from four separate data sets were analyzed. The composition of the study groups is displayed in Fig. 1.

1. Online samples 1a and 1b: EFA, CFA and ESEM on the CFQ and FSMC

The aim of the original study was to validate the German version of the 11-item CFQ and provide comparisons with the FSMC. Participants were recruited via advertisements at the website of the German Society for Multiple Sclerosis (DMSG) as well as the quarterly E-newsletter of the MS outpatient clinic, University Medical Centre Hamburg-Eppendorf. Inclusion criteria were self-reported MS diagnosis and age ≥ 18 . Exclusion criteria were major neurological or psychiatric comorbidities (except depression) as well as a MS relapse in the last 4 weeks. From the 873 individuals who registered on the website, 605 full data sets were obtained. For the purpose of the current analysis, the sample was split into two stratified halves (for details, see statistics section). All patients provided informed consent prior to enrolment. The local ethical review board approved the study (PV4772).

2. Online sample 2: CFA and ESEM on the CFQ and FSMC

For this sample, data were analyzed from the screening assessment for a randomized controlled trial of an online fatigue management program. Data of 293 PwMS were included in the present analyses. The inclusion criteria were a diagnosis of MS and age ≥ 18 . Exclusion criteria were major neurological or psychiatric comorbidities (except depression) as well as a MS relapse in the last 4 weeks. Participants were recruited via the website of the German MS Society. Out of the 293 patients, 120 did not only participate in the online program but were also diagnosed by a clinician and received a confirmed MS diagnosis. Confirmatory factor analyses were additionally performed for this group (Online Sample 2_{MS-diagnosis_confirmed}) to prove the representativeness of the results even if individuals participated in an online assessment. All patients provided informed consent prior to enrolment. The local ethical review board approved the study (PV4772).

3. Clinical sample 3: CFA and ESEM on the FSMC

This sample was obtained from the patient database of the MS outpatient clinic at the University Medical Center

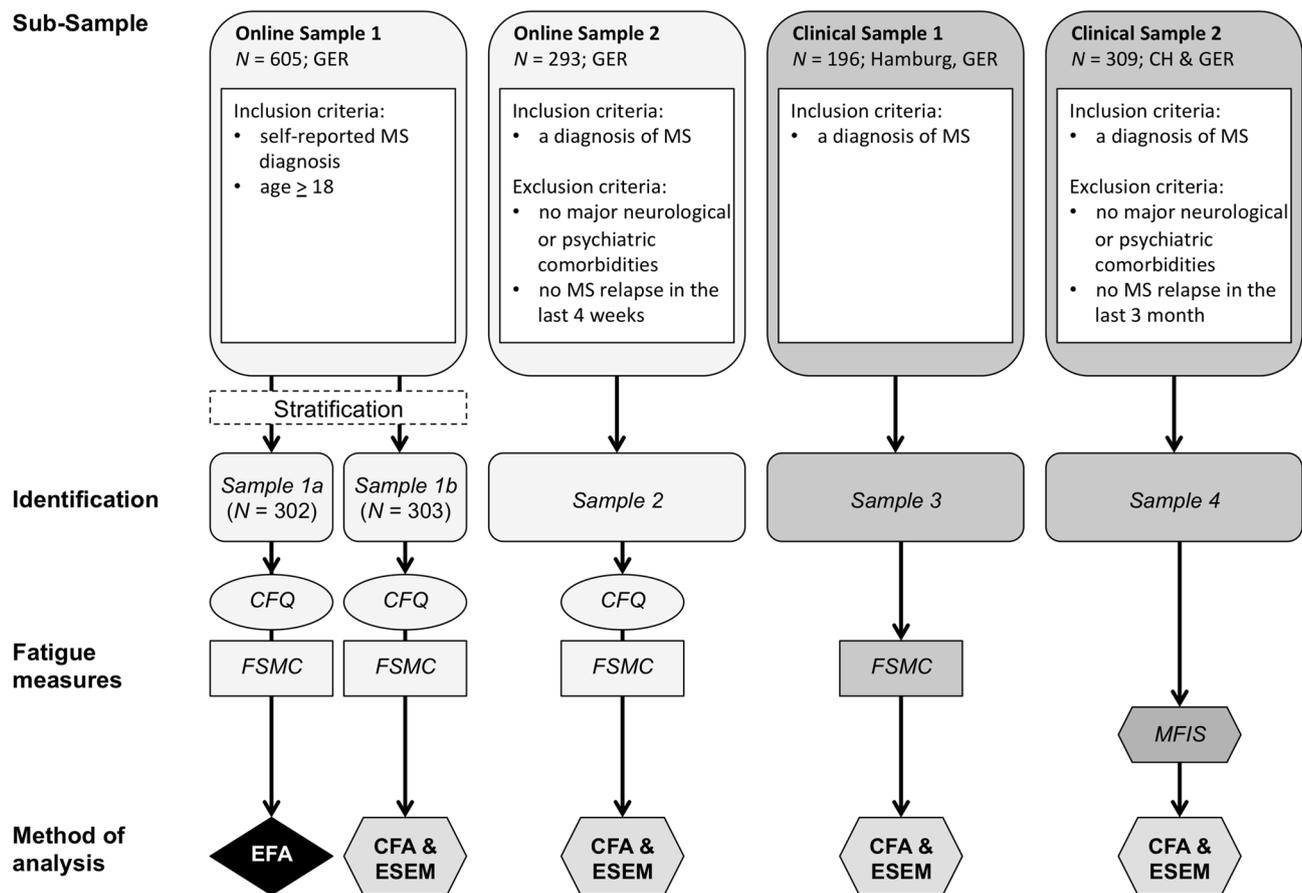


Fig. 1 Composition of the study samples and analytic strategy. *CFA* confirmatory factor analysis, *CFQ* Chalder Fatigue Questionnaire, *CH* Switzerland, *EFA* exploratory factor analysis, *ESEM* explora-

tory structural equation models, *FSMC* Fatigue Scale for Motor and Cognitive Functions, *GER* Germany, *MFIS* Modified Fatigue Impact Scale

Hamburg Eppendorf including consecutive PwMS who provided their clinical examination data by informed consent. All patients ($N = 196$) had a definite MS diagnosis according to revised diagnostic criteria [18]. PwMS were assessed in a single-test session, which took approximately 2.5 h. All patients completed the FSMC. The local ethical review board approved the study (PV4405).

4. Clinical sample 4: CFA and ESEM on the MFIS

For this sample, data were taken from the FSMC validation study, originally published by Penner et al. [10]. Data of 309 PwMS were utilized. Patients were eligible if they had definite MS according to McDonald's criteria and were without relapse during the last 3 months. Patients with a history of any other (non-MS-related) neurological and/or psychiatric disorder were excluded. All PwMS were outpatients recruited consecutively and face-to-face. All participants provided informed consent prior to enrolment.

The local ethics committee of the University of Basel, Switzerland approved the study.

Measures and procedures

Chalder Fatigue Questionnaire (CFQ)

The Chalder Fatigue Questionnaire (CFQ; [11]) contains 11 items referring to symptom severity in the past month. Responses are scored on a four-point scale (0–3). Item responses are added up for computation of (sub)scales. Three different scores can be computed: the global severity (CFQ_{total} ; all items; range 0–33), a physical fatigue sub-score ($CFQ_{physical}$; seven items; range 0–21) and a mental fatigue sub-score (CFQ_{mental} ; four items; range 0–12); see also [15]. This study used the German version of the CFQ after back-and-forward translations [19].

Fatigue Scale for Motor and Cognitive Functions (FSMC)

The Fatigue Scale for Motor and Cognitive Functions (FSMC) contains 20 items. Ten items measure symptoms of cognitive and motor fatigue, respectively [10]. Responses are scored on a five-point scale. Item scores can be summed to a global severity score (FSMC_{total}) or the two sub-scales (FSMC_{motor}; FSMC_{cognitive}) [10].

Modified Fatigue Impact Scale (MFIS)

The Modified Fatigue Impact Scale (MFIS) is the shortened version of the FIS [9]. The 21-item MFIS contains three sub-scales, a nine-item physical symptoms scale, a ten-item cognitive scale, and a two-item psychosocial scale. Items are scored on a five-point scale. Item scores can be summed to the three sub-scores (MFIS_{physical}; MFIS_{mental}; MFIS_{psychosocial}).

Hospital Anxiety and Depression Scale: German Version (HADS-D)

The Hospital Anxiety and Depression Scale—German Version (HADS-D) is a 14-item questionnaire, containing two sub-scales with seven items each [(1). Depression: HADS-D_{Depression}; (2) Anxiety: HADS-D_{Anxiety}; [20, 21]]. Items are rated on a four-point scale (0–3) and added for the computation of the sub-scale scores. For the present study, only the depression sub-scale was considered to account for the overlap between depression and fatigue in samples 1–3.

Beck Depression Inventory II

In sample 4, the Beck Depression Inventory (BDI-II, [22]) was used as a measure for depression.

Patient determined disease steps (PDDS)

Disability was assessed according to the disease steps defined by Hohol et al. [23] in samples 1 and 2, ranging from “no disability” (“0”) to “confined to wheelchair or worse” (“6”). Due to the presentation of the survey in the online surveys (1a, 1b, and 2), the participation of bedridden participants was not expected, so that the original version was preferred over the patient-determined disease steps (PDDS, NARCOMS) [24] scale, which suggests a further distinction of the severely impaired patients.

Expanded Disability Status Scale (EDSS)

The Expanded Disability Status Scale (EDSS; [25]) was applied as a physician-derived disability measure in samples 3 and 4, with disability scores ranging from “0” to “10”.

Statistical analyses

Demographic data of the samples were analyzed regarding differences in the specific sample characteristics. If the same measures were applied across different sub-samples, comparisons were made using Chi square tests or univariate analyses of variance (ANOVA) with Bonferroni-corrected pairwise comparisons. Inter-correlations between scales were calculated using Pearson correlations. All correlation coefficients and factor loadings could reach values in the range of -1 to $+1$. The magnitude in terms of effect size of the obtained correlation coefficients was interpreted according to Cohen [26] as “small” for coefficients ≥ 0.10 , “medium” for coefficients ≥ 0.30 and “large” for coefficients ≥ 0.50 .

For a cross-validation analysis based on sample 1, the sample was divided into two stratified samples (1a and 1b). Stratification can be used to randomly split a sufficiently large data set into two halves and derive an exploratory statistical model from one half that can be tested in the second half, while treating both halves as independent samples. Stratification criteria were derived from an initially calculated multivariate general linear model (GLM) with (1) the fixed factors sex, disease course and disability as well as (2) the respective two-way interactions with age as a covariate and (3) the CFQ and FSMC sub-scales as dependent variables. It aimed to identify potential confounders according to Brown [27]. Based on these considerations, we used the following stratification criteria: sex, disease course, and disability (as measured by the PDDS).

For the exploratory analysis of the underlying factor structure in the CFQ and FSMC, maximum likelihood factor analysis (MLFA) was calculated. Exploratory factor analysis estimates the best fitting factor structure in a given data set. The item-scale assignment thus solely depends on the data and does not follow a predefined theoretical model. In contrast, CFA is used to test whether the obtained data matches a predefined factor structure, independent from the given data set, i.e., the item-scale assignment defined in the initial publications of the scales. Deviations between the proposed factor structure and the obtained factor structure in the data are calculated and the model-fit determined. CFA and ESEM were both used to analyze model-fit. CFA has more rigid assumptions, i.e., that items loading onto one specific factor are not allowed to have cross-loadings with other hypothesized factors. Due to the rigid assumptions of CFA, it can fail to reach criteria for model-fit and lead to an overestimation of inter-factor correlations (for further details see [28–30]). To overcome this issue, more recent approaches suggest the use of exploratory structural equation models (ESEM) which provides a superior technique allowing free estimation of cross-loadings and utilizing oblique

Table 1 Demographic data—overview of the sub-sample characteristics

	Sample 1a	Sample 1b	Sample 2	Sample 3	Sample 4	Statistics
Number <i>N</i>	302	303	293	196	309	
Males %/females %	23.5%/76.5%	23.8%/76.2%	19.1%/80.2%	29.1%/70.9%	33.3%/66.6%	$\chi^2(4)=17.29, p=0.002$
Age						
Mean (SD)	41.9 (10.7)	41.8 (11.1)	41.4 (10.4)	40.9 (11.9)	43.4 (10.0)	$F(1396, 4)=2.13, p=0.075, \eta_p^2=0.006$
Disease course (%)						
Primary manifestation	2.3%	2.3%	–	1.5%	–	$\chi^2(8)=10.80, p=0.213$
RRMS	68.5%	68.6%	–	78.6%	–	
SPMS	15.9%	15.5%	–	9.7%	–	
PPMS	6.6%	6.9%	–	7.7%	–	
Not specified	6.6%	6.6%	–	2.6%	–	
PDDS						
Median	2	2	2	–	–	$\chi^2(14)=10.03, p=0.760$
EDSS						
Median	–	–	–	2.5	3.0	$\chi^2(16)=43.13, p<0.001$

RRMS relapsing–remitting multiple sclerosis, SPMS secondary progressive multiple sclerosis, PPMS primary progressive multiple sclerosis, EDSS Expanded Disability Status Scale, PDDS patient determined disease steps

rotation techniques [31]. This means that (1) items must show sufficiently large loadings on one factor but may have cross-loadings with other factors, and (2) that the factors themselves may be correlated. In the present study, we used ESEM in MPlus with the default oblique geomin rotation and the robust maximum likelihood estimation. MPlus was chosen as analysis tool as it provides a profound and sophisticated calculation algorithm for ESEMs and oblique rotation techniques that are well received in the international scientific community [32]. Cutoff criteria to evaluate the model-fit were defined according to Hu and Bentler [33], considering the thresholds for an at least acceptable fit: comparative fit index (CFI) ≥ 0.95 ; standardized root mean square residual (SRMR) < 0.10 ; non-normed fit index (NNFI) ≥ 0.95 ; root mean square error of approximation (RMSEA) ≤ 0.08 . The published thresholds for fit indices by Hu and Bentler are widely used (as evidenced by several thousand citations) and have also previously been applied in the confirmatory evaluation of fatigue scales, such as the work by Martin et al. [15] on the CFQ. As additional sensitivity analyses, CFA and ESEM were also computed for relevant subgroups, i.e., after exclusion of participants with clinically significant depressive symptoms ($HADS-D_{\text{Depression}} > 7/BDI-II > 9$) and in a sub-sample of the online cohort who had provided written neurologist confirmation of their MS diagnosis. Analyses were performed using MPlus 7.0 as well as R (including the Lavaan package). The level of significance was $\alpha = 0.05$.

Results

Sample characteristics

An overview of the sub-samples' demographic and clinical characteristics is provided in Table 1. Differences in demographic data were analyzed, if the respective variables were assessed in more than one sub-sample. The obtained sex difference resulted from a higher number of males in sample 4 that differed from the other four samples significantly [$\chi^2(1)=12.11, p=0.001$]. The sex effect vanished after removal of this sample [$\chi^2(3)=5.35, p=0.148$]. Sample 3 and 4 differed in the EDSS score. PwMS in sample 4 were on average more impaired.

Regarding the fatigue symptom severity (see Table 2), there were no significant difference between samples 1a, 1b, and 2 for the CFQ and FSMC (all Bonferroni-corrected post hoc pairwise comparisons $p > 0.05$). However, sample 3 differed from samples 1a, 1b, and 2 regarding the FSMC symptom severity; all Bonferroni-corrected post hoc tests were significant on a $p < 0.05$ level for the $FSMC_{\text{cognitive}}$ score and on a $p < 0.001$ level for the $FSMC_{\text{motor}}$ score, indicating lower symptom severity in sample 3. Since the MFIS was only assessed in sample 4 (physical: $M=19.6, SD=8.4$; mental: $M=16.0, SD=8.8$; psychosocial: $M=3.7, SD=2.2$), no comparisons could be calculated. Likewise, the categories for the disease course were different in samples 2 and 4. The group effect for differences in depression, as indicated by the $HADS-D_{\text{Depression}}$ score, was small; Bonferroni-corrected post hoc revealed no significant difference between any of the samples.

Table 2 Overview of fatigue scores in the sub-samples

	Sample 1a	Sample 1b	Sample 2	Sample 3	Sample 4	Statistics
CFQ _{physical}						
Mean (SD)	14.0 (3.5)	14.1 (3.4)	14.8 (3.5)	–	–	$F(895.2)=0.01, p=0.913 \eta_p^2<0.001$
CFQ _{mental}						
Mean (SD)	7.3 (2.3)	7.5 (2.3)	7.3 (2.3)	–	–	$F(895.2)=0.95, p=0.389 \eta_p^2=0.002$
FSMC _{motor}						
Mean (SD)	37.6 (7.1)	38.3 (7.6)	39.4 (6.1)	33.4 (8.3)	–	$F(1088, 3)=29.63, p<0.001 \eta_p^2=0.075$
FSMC _{cognitive}						
Mean (SD)	35.3 (8.0)	36.0 (8.6)	37.0 (7.5)	33.3 (8.7)	–	$F(1088, 3)=8.62, p<0.001 \eta_p^2=0.023$
HADS-D _{Depression}						
Mean (SD)	7.0 (3.9)	6.9 (3.8)	6.6 (3.8)	6.0 (4.1)	–	$F(1088, 3)=2.87, p<0.035 \eta_p^2=0.008$

CFQ Chalder Fatigue Questionnaire, FSMC Fatigue Scale for Motor and Cognitive Functions, HADS-D_{Depression} Hospital Anxiety and Depression Scale—German Version_subscale Depression

Generation of two stratified sub-samples

A multivariate GLM was calculated to identify factors associated with the CFQ- and FSMC-sub-scales for consideration in the stratification process. After exclusion of the non-significant two-way interactions and the non-significant main effect for age, three statistically significant main effects (sex: Wilk’s Lambda=0.98, $F(4, 585)=3.21, p=0.13$; disease steps: Wilk’s Lambda=0.78, $F(28, 2111)=5.46, p<0.001$; disease course: Wilk’s Lambda=0.94, $F(16, 1788)=3.17, p=0.001$) and one two-way interaction [sex × disease course: Wilk’s Lambda=0.95, $F(16, 1788)=3.30, p=0.020$] remained. The three factors were considered as stratification criteria. The assignment to the two sub-samples was randomized. After stratification, no significant differences were obtained in any of the variables samples 1a and 1b (mean p value: 0.98 ± 0.03).

Inter-correlations

The correlations between the FSMC and the CFQ as well as the sub-scales of these two fatigue measurements were statistically significant in all samples. Highest correlations with medium-to-large effect sizes were obtained between the two sub-scales of each of the two fatigue scales, i.e., FSMC_{motor} with FSMC_{cognitive} and CFQ_{mental} with CFQ_{physical}. Moreover, all fatigue scales showed significant correlations with the HADS-D_{Depression} with medium effect sizes, indicating a conceptual overlap between the constructs (Table 3). Likewise, medium effect sizes were obtained between opposing sub-scales in the two different fatigue measures FSMC and CFQ.

Table 3 Inter-correlations between the HADS-D_{Depression} and the fatigue measures CFQ and FSMC across the sub-samples 1a, 1b, 2, and 3

	HADS-D _{Depression}	FSMC _{cognitive}	FSMC _{motor}	CFQ _{mental}
FSMC _{cognitive}	(1a) 0.29** (1b) 0.34** (2) 0.35** (3) 0.34**			
FSMC _{motor}	(1a) 0.24** (1b) 0.29** (2) 0.38** (3) 0.37**	(1a) 0.65** (1b) 0.75** (2) 0.54** (3) 0.70**		
CFQ _{mental}	(1a) 0.25** (1b) 0.29** (2) 0.18* (3) –	(1a) 0.53** (1b) 0.65** (2) 0.55** (3) –	(1a) 0.22** (1b) 0.43** (2) 0.24** (3) –	
CFQ _{physical}	(1a) 0.28** (1b) 0.35** (2) 0.28** (3) –	(1a) 0.39** (1b) 0.51** (2) 0.29** (3) –	(1a) 0.47** (1b) 0.56** (2) 0.34** (3) –	(1a) 0.46** (1b) 0.53** (2) 0.54** (3) –

CFQ Chalder Fatigue Questionnaire, FSMC Fatigue Scale for Motor and Cognitive Functions, HADS-D_{Depression} Hospital Anxiety and Depression Scale—German Version_subscale Depression

** $p<0.001$; * $p<0.01$. As in sample 3, the CFQ was not administered, these numbers are left blank “–”

Chalder Fatigue Questionnaire: exploratory factor analysis

The factor structure of the CFQ was analyzed using MLFA (maximum likelihood axis factoring, VARIMAX-rotation), based on data of sample 1a. The screen-plot as well as the analysis of the Eigenvalues favored the extraction of two factors. Factor loadings for the rotated factor solution are displayed in Table 4. Both factors account for 48.1% of the

Table 4 Results for the exploratory factor analyses for the Chalder Fatigue Questionnaire based on a two-factor solution (sample 1a)

Item	Two-factor solution	
	Factor 1	Factor 2
Subscale CFQ _{physical}		
01. CFQ _{physical}	0.71	0.21
02. CFQ _{physical}	0.69	0.26
03. CFQ _{physical}	0.71	0.22
04. CFQ _{physical}	0.55	0.21
05. CFQ _{physical}	0.67	0.20
06. CFQ _{physical}	0.44	0.08
07. CFQ _{physical}	0.64	0.14
Subscale CFQ _{mental}		
08. CFQ _{mental}	0.36	0.54
09. CFQ _{mental}	0.20	0.81
10. CFQ _{mental}	0.15	0.82
11. CFQ _{mental}	0.17	0.59
Variance explained	27.7%	20.4%

Numbers display the individual factor loadings for all items according to a two-way solution after VARIMAX-rotation

CFQ Chalder Fatigue Questionnaire

scale variance. The assignment of the scale items to the two factors matched the originally proposed factor structure.

FSMC: exploratory factor analysis

The factor structure of the FSMC was analyzed using the same methodological approach as described for the CFQ (MLFA-maximum likelihood axis factoring, VARIMAX-rotation). The factor loadings of the EFA are displayed in Table 5. Both factors account for 46.5% of the scale variance. The assignment of the scale items to the two factors matched the originally proposed factor structure from the original validation study for most items [10]. Relevant cross-loadings were obtained in four of the 20 items.

Confirmatory factor analysis for CFQ, FSMC, and MFIS

CFA was used to test the model-fit of the proposed factor structure of the CFQ, based on data of samples 1b, 2, and 3. The data did not confirm the proposed factor structure. Fit indices did not improve after exclusion of participants with clinically relevant depression symptoms. CFA were also calculated for the FSMC, based on the data of samples 1b, 2, and 3. Similar to the results obtained for the CFQ, the hypothesized model had to be rejected. Again, exclusion of participants with significant depression did not improve the model-fit. For the MFIS (sample 4), none of the calculated

Table 5 Results for the exploratory factor analyses for the FSMC, based on two-factor solution (sample 1a)

Item	Two-factor solution	
	Factor 1	Factor 2
Subscale FSMC _{Cognitive}		
01. FSMC _{cognitive}	0.60	0.24
04. FSMC _{cognitive}	0.62	0.19
07. FSMC _{cognitive}	0.67	0.19
08. FSMC _{cognitive}	0.56	0.21
11. FSMC _{cognitive}	0.70	0.22
13. FSMC _{cognitive}	0.39	0.43
15. FSMC _{cognitive}	0.62	0.18
17. FSMC _{cognitive}	0.74	0.13
18. FSMC _{cognitive}	0.81	0.21
20. FSMC _{cognitive}	0.75	0.13
Subscale FSMC _{Motor}		
02. FSMC _{motor}	0.25	0.70
03. FSMC _{motor}	0.27	0.65
05. FSMC _{motor}	0.48	0.45
06. FSMC _{motor}	0.38	0.31
09. FSMC _{motor}	0.04	0.71
10. FSMC _{motor}	0.13	0.76
12. FSMC _{motor}	0.37	0.48
14. FSMC _{motor}	0.24	0.70
16. FSMC _{motor}	0.63	0.35
19. FSMC _{motor}	0.16	0.48
Variance explained	27.1%	19.3%

The FSMC items keep their original item numbers. Numbers display the individual factor loadings for all items. According to a two-way solution after VARIMAX-rotation

FSMC Fatigue Scale for Motor and Cognitive Functions

CFA matched the fit criteria either. Results are displayed in Table 6.

Exploratory structural equation modeling for CFQ, FSMC, and MFIS

ESEMs were calculated for the CFQ, the FSMC and the MFIS, with a hypothesized 2-factor structure for the CFQ and FSMC as well as a 3-factor structure for the MFIS. As for the CFA, none of the calculated ESEM led to a sufficient model-fit, even though this method allows cross-loadings as well as oblique rotation (Table 7).

Discussion

The aim of the present study was to analyze the factorial statistical validity of three widely used fatigue instruments in MS (CFQ, FSMC, and MFIS) to evaluate the

Table 6 Results of the CFA, calculated for the CFQ, the FSMC, and the MFIS

	χ^2	<i>p</i>	CMIN/ <i>df</i>	CFI	NNFI	RMSEA
Chalder Fatigue Questionnaire (CFQ)						
Total sample						
1b (<i>N</i> =303)	234.6	<0.001	5.46	0.85	0.80	0.12
2 (<i>N</i> =293)	190.9	<0.001	4.44	0.89	0.85	0.11
2 _{MS-diagnosis_confirmed} (<i>N</i> =120)	121.9	<0.001	2.84	0.86	0.82	0.12
Without participants HADS- <i>D</i> _{Depression} > 7						
1b (<i>N</i> =182)	131.4	<0.001	3.06	0.88	0.84	0.11
2 (<i>N</i> =180)	117.4	<0.001	2.73	0.89	0.86	0.10
Fatigue Scale for Motor and Cognitive Functions (FSMC)						
Total sample						
1b (<i>N</i> =303)	901.1	<0.001	5.33	0.81	0.78	0.12
2 (<i>N</i> =293)	730.8	<0.001	4.32	0.77	0.74	0.11
2 _{MS-diagnosis_confirmed} (<i>N</i> =120)	481.3	<0.001	2.85	0.71	0.67	0.12
3 (<i>N</i> =196)	540.3	<0.001	3.20	0.84	0.82	0.11
Without participants HADS- <i>D</i> _{Depression} > 7						
1b (<i>N</i> =182)	644.1	<0.001	3.81	0.80	0.77	0.13
2 (<i>N</i> =180)	560.15	<0.001	3.31	0.72	0.69	0.11
3 (<i>N</i> =134)	429.5	<0.001	3.16	0.84	0.83	0.11
Modified Fatigue Impact Scale (MFIS)						
Total sample						
4 (<i>N</i> =309)	603.93	<0.001	3.25	0.93	0.92	0.09
Without participants BDI-II > 9						
4 (<i>N</i> =176)	454.91	<0.001	2.45	0.91	0.90	0.09

BDI-II Beck Depression Inventory II, *CFI* comparative fit index, *HADS-*D*_{Depression}* Hospital Anxiety and Depression Scale—German Version, *NNFI* Nonnormed Fit Index, *RMSEA* root mean square error of approximation

appropriateness of the proposed theoretical fatigue dimensions. According to the EFA, the CFQ and FSMC largely matched the originally published factor structure from the validation studies. However, for the FSMC, the resulting factor loadings for some items were ambiguous (items 5, 6, 13, and 16). CFA and ESEM showed that the proposed factor structure did not fit the data for any of the evaluated fatigue instruments. Even the exclusion of patients with depressive symptoms did not improve the model-fit. This suggests that the findings are unlikely due to a potential overlap of depression and fatigue. The use of ESEMs with oblique rotation also did not lead to a sufficient model-fit. Thus, from a statistical point of view, our results challenge the proposed fatigue dichotomy in its present form by means of self-rating instruments.

The large diversity in assessment tools for fatigue demonstrates that there is still no consensus on the nature of fatigue as well as its reliable and valid assessment [10]. Researchers tend to agree on the common distinction between cognitive and motor fatigue [34], which mainly stems from clinical observations and patients' self-reports. The CFQ and FSMC are both widely applied measures and are currently among the best validated measures in

the field [35]. However, studies that focused on the re-evaluation of scales such as the CFQ failed to prove a valid distinction of the original scales by means of CFA [15, 16, 36].

From a statistical viewpoint, the results are twofold: the poor model-fit in the CFA should not result in what is called the Henny-Penny-Problem [37], i.e., that researchers call into question all previous results based on poor fit indices. CFA in particular can generate a model misfit when minor items exhibit residual covariation, for example due to non-specified underlying factors [38]. Therefore, ESEM was introduced as a new method to overcome the rigid assumptions of CFA. In their data-driven approach, Fong et al. [36] already applied an ESEM approach to the CFQ and could not replicate the originally proposed dichotomy. Instead, an adaptation of the factor structure was necessary to achieve an acceptable model-fit. However, a utilization of the exploratory nature of such purely data-driven approaches to verify theoretical models that are not supported by data should be avoided. Instead, theoretical considerations and research findings from other levels of analysis must be acknowledged and should be integrated into a sound scientific framework and conceptualization of fatigue.

Table 7 Results of the ESEMs, calculated on the CFQ, the FSMC and the MFIS

	χ^2	<i>p</i>	CMIN/ <i>df</i>	CFI	SRMR	RMSEA (90% CI)
Chalder Fatigue Questionnaire (CFQ)						
Total sample						
1b (<i>N</i> = 303)	135.31	<0.001	3.98	0.92	0.04	0.10 (0.08–0.12)
2 (<i>N</i> = 293)	129.39	<0.001	3.81	0.93	0.05	0.10 (0.08–0.12)
2 _{MS-diagnosis_confirmed} (<i>N</i> = 120)	101.99	<0.001	2.27	0.90	0.06	0.10 (0.08–0.13)
Without participants HADS- <i>D</i> _{Depression} > 7						
1b (<i>N</i> = 182)	82.89	<0.001	2.44	0.93	0.05	0.09 (0.07–0.11)
2 (<i>N</i> = 180)	88.63	<0.001	2.61	0.92	0.05	0.09 (0.07–0.12)
Fatigue Scale for Motor and Cognitive Functions (FSMC)						
Total sample						
1b (<i>N</i> = 303)	707.16	<0.001	4.68	0.86	0.05	0.11 (0.10–0.12)
2 (<i>N</i> = 293)	582.27	<0.001	3.90	0.82	0.06	0.10 (0.09–0.11)
2 _{MS-diagnosis_confirmed} (<i>N</i> = 120)	438.97	<0.001	2.57	0.75	0.08	0.11 (0.10–0.13)
3 (<i>N</i> = 196)	389.98	<0.001	2.58	0.89	0.05	0.09 (0.08–0.10)
Without participants HADS- <i>D</i> _{Depression} > 7						
1b (<i>N</i> = 182)	514.36	<0.001	3.41	0.85	0.06	0.12 (0.10–0.13)
2 (<i>N</i> = 180)	463.20	<0.001	3.07	0.78	0.07	0.11 (0.10–0.12)
3 (<i>N</i> = 134)	333.29	<0.001	2.21	0.89	0.05	0.09 (0.08–0.11)
Modified Fatigue Impact Scale (MFIS)						
Total sample						
4 (<i>N</i> = 309)	362.09	<0.001	2.41	0.96	0.02	0.07 (0.06–0.08)
Without participants BDI-II > 9						
4 (<i>N</i> = 176)	266.03	<0.001	1.77	0.96	0.03	0.07 (0.05–0.08)

BDI-II Beck Depression Inventory II, *CFI* comparative fit index, *CI* confidence interval, *HADS-D*_{Depression} Hospital Anxiety and Depression Scale—German Version_subscale Depression, *RMSEA* root mean square error of approximation, *SRMR* Standardized Root Mean Square Residual

Some limitations might require consideration. The MS diagnoses of some of the participants from the online assessments were not confirmed by a physician, as the study was performed via the Internet. Due to the online sampling method, there could have also been a selection bias. However, we re-ran our analyses in a sub-sample of 120 patients who had provided written confirmation of MS diagnosis by their neurologist and obtained the same result. Thus, while self-reported diagnosis of MS might not always be reliable, our data strongly suggest that this is unlikely to explain our results. Moreover, data on disease course are missing for the two online cohorts and lower reliability of the PDDS particularly at the lower end of the scale must be considered. Additionally, different instructions and time frames considered in the different measures further complicate the comparison of the different CFQs. As the fit measures for the MFIS and CFQ were slightly better than the once for the FSMC, the larger time frame used in the FSMC could have accounted for a less precise retrieval of information of past fatigue symptoms.

The purpose of our analyses was to explore whether the MS-fatigue questionnaires, in their currently available form, do or do not justify the use of presumably distinguishable

subconstructs of MS-related fatigue. Based on multiple analyses in several large cohorts of MS patients, the answer appears to be “no”. The present analyses thus highlight the necessity to refine existing measures to meet psychometric requirements. From a clinical perspective, an unequivocal assessment of fatigue in PwMS is an essential prerequisite for further research, e.g., into putative pathobiological substrates of fatigue (sub-)types as well as outcomes for treatment interventions. The development of interventions for the reduction of fatigue should therefore take into consideration potential shortcomings in the current definitions of fatigue dimensions and therefore focus on total fatigue scores than on fatigue sub-scores.

Acknowledgements We thank Lena Katharina Feddersen for support with data collection.

Compliance with ethical standards

Conflicts of interest GEAP received speaker honoraria and project funding from Genzyme Sanofi. CH received speaker honoraria and project funding from Genzyme Sanofi, Roche, Merck, Biogen. SMG reports honoraria from Mylan GmbH, Almirall S. A. and Celgene and research grants from Biogen. IKP has received honoraria for speaking

at scientific meetings, serving at scientific advisory boards and consulting activities from Adamas Pharma, Almirall, Bayer Pharma, Biogen, Celgene, Desitin, Genzyme, Merck, Novartis, Roche, and Teva. She has received research support from Teva, Novartis and the German MS Society. All remaining authors have no conflict of interest to declare.

Ethical standards All studies were conducted in accordance with the 1964 Declaration of Helsinki and its later amendments and received ethical clearance.

Informed consent All persons provided written informed consent prior to their inclusion in the study.

References

1. Penner IK, Paul F (2017) Fatigue as a symptom or comorbidity of neurological diseases. *Nat Rev Neurol*. <https://doi.org/10.1038/nrneurol.2017.117>
2. Barin L, Salmen A, Disanto G et al (2018) The disease burden of multiple sclerosis from the individual and population perspective: which symptoms matter most? *Mult Scler Relat Disord*. <https://doi.org/10.1016/j.msard.2018.07.013>
3. Gerhard L, Dorstyn DS, Murphy G, Roberts RM (2018) Neurological, physical and sociodemographic correlates of employment in multiple sclerosis: a meta-analysis. *J Health Psychol*. <https://doi.org/10.1177/1359105318755262>
4. Kobelt G, Thompson A, Berg J, Gannedahl M, Eriksson J, MSCOI Study Group, & European Multiple Sclerosis Platform (2017) New insights into the burden and costs of multiple sclerosis in Europe. *Mult Scler J*. <https://doi.org/10.1177/1352458517694432>
5. Vister E, Tijmsa ME, Hoang PD, Lord SR (2017) Fatigue, physical activity, quality of life, and fall risk in people with multiple sclerosis. *Int J MS Care*. <https://doi.org/10.7224/1537-2073.2015-077>
6. Paucke M, Kern S, Ziemssen T (2018) Fatigue and sleep in multiple sclerosis patients: a comparison of self-report and performance-based measures. *Front Neurol*. <https://doi.org/10.3389/fneur.2017.00703>
7. Kos D, Kerckhofs E, Nagels G, D'hooghe MB, Ilsbrouckx S (2008) Origin of fatigue in multiple sclerosis: review of the literature. *Neurorehabilit Neural Repair*. <https://doi.org/10.1177/1545968306298934>
8. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD (1989) The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol*. <https://doi.org/10.1001/archneur.1989.00520460115022>
9. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ (1994) The impact of fatigue on patients with multiple sclerosis. *Can J Neurol Sci*. <https://doi.org/10.1017/S0317167100048691>
10. Penner IK, Raselli C, Stöcklin M, Opwis K, Kappos L, Calabrese P (2009) The Chalder fatigue questionnaire for motor and cognitive functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. *Mult Scler J*. <https://doi.org/10.1177/1352458509348519>
11. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, Wallace EP (1993) Development of a fatigue scale. *J Psychosom Res*. [https://doi.org/10.1016/0022-3999\(93\)90081-P](https://doi.org/10.1016/0022-3999(93)90081-P)
12. Fischer JS, LaRocca NG, Miller DM et al (1999) Recent developments in the assessment of quality of life in multiple sclerosis (MS). *Mult Scler J*. <https://doi.org/10.1177/135245859900500410>
13. Larson RD (2013) Psychometric properties of the modified fatigue impact scale. *Int J MS Care*. <https://doi.org/10.7224/1537-2073.2012-019>
14. Mills RJ, Young CA, Pallant JF, Tennant A (2010) Rasch analysis of the modified fatigue impact scale (MFIS) in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. <https://doi.org/10.1136/jnnp.2008.151340>
15. Martin A, Staufenbiel T, Gaab J, Rief W, Brähler E (2010) Messung chronischer Erschöpfung-Teststatistische Prüfung der Fatigue Skala (FS). *Zeitschrift für Klinische Psychologie und Psychotherapie*. <https://doi.org/10.1026/1616-3443/a000010>
16. Chilcot J, Norton S, Kelly ME, Moss-Morris R (2016) The Chalder Fatigue Questionnaire is a valid and reliable measure of perceived fatigue severity in multiple sclerosis. *Mult Scler J*. <https://doi.org/10.1177/1352458515598019>
17. Picariello F, Moss-Morris R, Macdougall IC, Chilcot J (2016) Measuring fatigue in haemodialysis patients: the factor structure of the Chalder Fatigue Questionnaire (CFQ). *J Psychosom Res*. <https://doi.org/10.1016/j.jpsychores.2016.03.124>
18. Polman CH, Reingold SC, Banwell B et al (2011) Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol*. <https://doi.org/10.1002/ana.22366>
19. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186–3191
20. Herrmann C, Buss U, Snaith RP (1995) HADS-D hospital anxiety and depression scale—deutsche version. Verlag Hans Huber, Bern
21. Petermann F (2011) Hospital anxiety and depression scale. Deutsche Version (HADS-D). *Zeitschrift für Psychiatrie Psychologie und Psychotherapie*. <https://doi.org/10.1024/1661-4747/a000077>
22. Hautzinger M, Keller F, Kühner C (2006) BDI-II beck depressions-inventar: revision. Harcourt Test Services, Frankfurt am Main
23. Hohol MJ, Orav EJ, Weiner HL (1995) Disease steps in multiple sclerosis. A simple approach to evaluate disease progression. *Neurology*. <https://doi.org/10.1212/WNL.45.2.251>
24. Learmonth YC, Motl RW, Sandroff BM, Pula JH, Cadavid D (2013) Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis. *BMC Neurol*. <https://doi.org/10.1186/1471-2377-13-37>
25. Kurtzke JF (1983) Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. <https://doi.org/10.1212/WNL.33.11.1444>
26. Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd edn. Erlbaum, Hillsdale
27. Brown TA (2014) *Confirmatory factor analysis for applied research*. Guilford Publications, New York
28. Asparouhov T, Muthén B (2009) Exploratory structural equation modeling. *Struct Equ Model Multidiscip J*. <https://doi.org/10.1080/10705510903008204>
29. Marsh HW, Lüdtke O, Muthén B, Asparouhov T, Morin AJ, Trautwein U, Nagengast B (2010) A new look at the big five factor structure through exploratory structural equation modeling. *Psychol Assess*. <https://doi.org/10.1037/a0019227>
30. Marsh HW (2007) Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology. *Handb Sport Psychol*. <https://doi.org/10.1002/9781118270011.ch35>
31. Morin AJ, Marsh HW, Nagengast B (2013) Exploratory structural equation modeling. In: Hancock GR, Mueller RO (eds) *Quantitative methods in education and the behavioral sciences: issues, research, and teaching*. Structural equation modeling: a second course. IAP Information Age Publishing, Charlotte, pp 395–436
32. Byrne BM (2016) *Structural equation modeling with Mplus: basic concepts, applications, and programming*. Routledge, New York
33. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus

- new alternatives. *Struct Equ Model Multidiscip J*. <https://doi.org/10.1080/10705519909540118>
34. Claros-Salinas D, Bratzke D, Greitemann G, Nickisch N, Ochs L, Schröter H (2010) Fatigue-related diurnal variations of cognitive performance in multiple sclerosis and stroke patients. *J Neurol Sci*. <https://doi.org/10.1016/j.jns.2010.04.018>
35. Elbers RG, Rietberg MB, van Wegen EE, Verhoef J, Kramer SF, Terwee CB, Kwakkel G (2011) Self-report fatigue questionnaires in multiple sclerosis Parkinson's disease and stroke: a systematic review of measurement properties. *Qual Life Res* 21(6):925–944. <https://doi.org/10.1007/s11136-011-0009-2>
36. Fong TC, Chan JS, Chan CL et al (2015) Psychometric properties of the Chalder Fatigue Questionnaire revisited: an exploratory structural equation modeling approach. *Qual Life Res*. <https://doi.org/10.1007/s11136-015-0944-4>
37. Hopwood CJ, Donnellan MB (2010) How should the internal structure of personality inventories be evaluated? *Personal Soc Psychol Rev*. <https://doi.org/10.1177/1088868310361240>
38. Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJ, Trautwein U (2009) Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Struct Equ Model Multidiscip J*. <https://doi.org/10.1080/10705510903008220>