Clinical pain research

Egil A. Fors*, Knut-Arne Wensaas, Hilde Eide, Ellen A. Jaatun, Daniel J. Clauw, Fred Wolfe and Anne-Sofie Helvik

Fibromyalgia 2016 criteria and assessments: comprehensive validation in a Norwegian population

<https://doi.org/10.1515/sjpain-2020-0002>

Received December 20, 2019; revised May 3, 2020; accepted May 5, 2020

Abstract

Background and aims: The ACR1990 criteria of fibromyalgia (FM) have been criticized due to poor reliability of tender points counting (TPC), inconsistent definitions of the widespread pain, and by not considering other symptoms than pain in the FM phenotype. Therefore, several newer self-report measures for FM criteria have emerged. The aim of this study was to translate the fibromyalgia survey questionnaire (FSQ) to Norwegian and validate both the 2011 and the 2016 fibromyalgia survey diagnostic criteria (FSDC) against the ACR1990 criteria.

***Corresponding author: Egil A. Fors**, Faculty of Medicine and Health Sciences, Department of Public Health and Nursing, General Practitioner Research Unit, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway, Phone: +4741236597, E-mail: egil.a.fors@ntnu.no

Knut-Arne Wensaas: Research Unit for General Practice, NORCE Norwegian Research Centre, Bergen, Norway

Hilde Eide: Science Centre Health and Technology, Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

Ellen A. Jaatun: Norwegian University of Science and Technology, Trondheim, Norway; and St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

Daniel J. Clauw: Chronic Pain and Fatigue Research Center, University of Michigan Health System, Ann Arbor, Michigan, USA; and Department of Anesthesiology, University of Michigan Health System, Ann Arbor, Michigan, USA

Fred Wolfe: Forward, the National Data Bank for Rheumatic Diseases, Wichita, KS, USA; and University of Kansas School of Medicine, Wichita, KS, USA

Anne-Sofie Helvik: Faculty of Medicine and Health Sciences, Department of Public Health and Nursing, General Practitioner Research Unit, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway

Methods: One hundred and twenty chronic pain patients formerly diagnosed with fibromyalgia according to the ACR1990 criteria, and 62 controls not diagnosed or where fibromyalgia was not suspected, were enrolled in this study. All responded to a Norwegian version of the FSQ. Also, they had a clinical examination according to ACR1990 fibromyalgia criteria including a counting of significant tender points with an algometer (TPC). The FSQ with the Widespread Pain Index (WPI) and Symptom Severity scale (SSS) subscales, Fibromyalgia Severity (FS) sum score, was examined for correlations with the fibromyalgia impact questionnaire (FIQ) and TPCs. Face-validity, internal consistence, test-retest reliability and construct validity with convergent and divergent approaches were examined and a Receiver Operating Characteristics (ROC) analysis was performed.

Results: The internal consistency of FS measured by Cronbach's α was good ($=0.904$). The test-retest reliability measures using intra class correlation were respectable for the FS, including WPI and SSS subscales (0.86, 0.84 and 0.87). FS, WPI and SSS correlated significantly with FIQ (0.74, 0.59 and 0.85) and TPC indicating an adequate construct, convergent validity. The medians of FS, WPI and SSS in the fibromyalgia-group were significantly different from the non-fibromyalgia-group indicating good construct, divergent validity. Using the 2011 and 2016 FSDC vs. ACR 1990 as a reference, sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR-) were identified. The accuracy rate for both 2011 and 2016 FSDC were respectable (84%). ROC analysis using FS revealed a very good Area Under the Curve (AUC) = 0.860.

Conclusion: The current study revealed that the Norwegian versions of FSQ is a valid tool for assessment of fibromyalgia according to the 2011 and 2016 (FSDC).

Keywords: fibromyalgia; 2016 survey and diagnostic criteria; validity; Norwegian; questionnaire.

1 Introduction

The diagnosis of fibromyalgia (FM) is based on criteria set by the American College of Rheumatology (ACR) [1]. In 2010 new provisional ACR criteria for diagnosing FM and measuring symptom severity were published [2], as a result of a series of objections to the former criteria [1]. The 1990 criteria defined FM based on chronic widespread pain (CWP) and tenderness in 11 out of 18 predefined tender points (TP) [1], but were later criticized in both clinical practice [3] and research [2, 4].

Many physicians did not know how to examine for tender points or refused to do it, and consequently, a correct evaluation of tender points was rarely performed [5, 6]. In addition, tender points are often interpreted as muscle (peripheral) pathology, while contemporary research suggest that changes in central nervous mechanisms are the likely pathogenic processes [7]. Another limitation of the ACR 1990 criteria is that they do not consider symptoms beyond pain, even though it is clear that the fibromyalgia phenotype often expresses other symptoms like fatigue, sleep disorders, cognitive dysfunction, and somatic symptoms [2], recognized previously [8, 9]. An additional problem is that the ACR 1990 criteria do not grade the severity of FM symptoms or monitor quantitative changes in FM outcomes since they were simply dichotomous, i.e. “whether you have fibromyalgia or not”. Thus, researchers have criticized the ACR1990 criteria [7, 10–13] and raised questions as whether «fibromyalgia falls foul of a fallacy» [14]. The shift in the conceptualization of fibromyalgia that occurred in the clinic and in research studies, however, provided no clear new case definition. The new ACR 2010 FM diagnostic criteria (FDC) shifted the fibromyalgia definition somewhat toward other important symptoms, not only pain. Thus, the 2010 criteria were not aiming for perfect agreement with 1990 criteria because the FM definition was changed to some extent [2].

The ACR 2010 FM diagnostic criteria were based on a Widespread Pain Index (WPI) (range 0–19) and a Symptom Severity Scale (SSS) score (range 0–12). One of the items of the SSS required an assessment from a physician, and in 2011 the criteria were modified so that all items could be obtained by patient self-administration using the Fibromyalgia Survey Questionnaire (FSQ) for research [15]. The 2011 Fibromyalgia survey diagnostic criteria (FSDC) were satisfied if the following three conditions are met: (1) the WPI ≥ 7 and the SSS score ≥ 5 , or WPI is 3–6 plus SSS score ≥ 9 ; (2) symptoms have been present at a similar level for at least 3 months; and (3) the patient does not have a disorder that would otherwise explain the pain [15].

Based on research published in 2010–2016 the FSDC were updated in 2016. These criteria were evaluated to adequately serve as diagnostic criteria when used in the clinic, but *also* as classification criteria when used for research [4]. The 2016 criteria are satisfied if the following three conditions are met: (1) the WPI ≥ 7 and the SSS score ≥ 5 , or WPI is 4–6 with SSS score ≥ 9 ; (2) generalized pain, defined as pain in at least four of five regions, is present; and (3) symptoms have been present at a similar level for at least 3 months. The previous requirement that the patient could not have other conditions that could explain the pain was removed (4). The sum of the WPI and the SSS is termed the Fibromyalgia Severity (FS) score (range 0–31), but also the term Fibromyalgianess score has been used [2, 4, 15, 16]. The FSQ has been translated into some languages, including German [17], Spanish [18], Persian [19], Japanese [20], Turkish [21], French [22] and Korean [23], and validated in these populations. The aim of this project is to validate the Norwegian version of FSQ according to the 2011 and 2016 FSDC against the ACR 1990 FM criteria.

2 Methods

2.1 Design and participants

Participants were enrolled from September 2016 to September 2019. Patients with established or newly diagnosed fibromyalgia according to the ACR1990 criteria were eligible for the FM case group and invited by the Norwegian Fibromyalgia Association (NFA) or the Norwegian Rheumatism Association (NRA) to participate in the study. The patient associations invited the participants by mail, telephone or personal contacts. The control group was recruited through three different channels: (1) the NFA administered website where non-FM diagnosed visitors were asked to participate, (2) non-FM patients with rheumatic conditions such as rheumatoid arthritis and osteoarthritis recruited by the NRA, and (3) employees at two outpatient clinics (Coperio Medical Center and a university hospital otorhinolaryngology clinic) using so-called “snowball subject recruitment” [24]. Allocation to the FM-case or control groups was based on the response to the fibromyalgia screening question in the HUNT3 survey: “Have you had, or do you have any of the following: Fibromyalgia”? ... (among other conditions, e.g. low back pain, etc.) with response alternatives “yes” or “no” (HUNT 3 Q1) [25]. In addition, they were asked: ... or “are you

under consideration for a FM diagnosis?”. The participants were also asked for a “history of widespread pain” for at least 6 months and made a body pain drawing to reveal pain in 4 quadrants and the axial region, necessary for a fibromyalgia diagnosis. Participants <18 and >70 years of age were excluded. All participants signed an informed consent, either electronically or on paper, before inclusion. The study was approved by the Regional Committee for Medical and Health Research Ethics (project REK Nord # 2014/938), including fulfillment of the General Data Protection Regulation and in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2 Translation procedure

Two Norwegian pain researchers fluent in English (HE + EF) independently translated the FSQ from English to Norwegian and then harmonized the two translations. A consolidation of the Norwegian version was performed by 6 FM patients and three other Norwegian pain researchers, who found the content adequate and understandable. Thereafter, this Norwegian version was translated back into English by a professional translation agency. This back-translation was checked by authors of the original English version (FW, DC and team). The Norwegian version of the FSQ (FSQ-N) was then accepted for further studies.

2.3 Assessments

All participants answered questions about their sociodemographic status (age, marital status and habitancy, economy, employment and financial status), psychological measures (anxiety and depression) [26] and pain intensity (VAS 0–10). They completed the Norwegian version of the FSQ and the Fibromyalgia Impact Questionnaire (FIQ). A subgroup was asked to re-score the Fibromyalgia Survey Questionnaire after 1 week. All questionnaires were returned by mail or an electric survey platform hosted by NFA.

The *FIQ* is a questionnaire often used to examine the impact of FM on function and level of symptoms during the previous week [27, 28]. It has 10 items (20 questions) and measures function, overall/work impact, symptom score and FIQ total score. The FIQ is scored in such a way that a higher score indicates a greater impact of the syndrome on

the person. The FIQ function (physical impairment) is calculated from 11 questions related to the ability to perform day-to-day activities scored on a 4-point Likert scale from “always” (0) to “never” (3). The item scores are summed, then divided by the numbers of completed questions and multiplied by 3.33, to give an FIQ function score (range 0–10, where a higher score represents more physical impairment). The FIQ overall/work impact is calculated from the number of days with overall well-being and the impact on working ability (on an 8-point scale from no (0) to maximum (7) impact) during the previous week (range 0–20, where a higher score represents greater negative impact). The FIQ symptom score is the sum of the severity (rated from “no symptoms” (0) to “substantial symptoms” (10) of 7 symptoms (range 0–70)). The FIQ *total* is the sum score of all items (range 0–100). The average fibromyalgia patient will have an FIQ total score of about 50; severely afflicted patients usually score 70 or higher [28].

2.3.1 Tender points counting and widespread pain

The participants were examined and classified as having FM or not according to the ACR1990 criteria [1]. In addition to responding whether they had a history of widespread pain >6 months, they completed an electronic pain drawing [29] in order to detect pain in 4 quadrants and axial, i.e. pain in the left side of the body, pain in the right side of the body, pain above the waist, pain below the waist and axial, skeletal pain. Thereafter, pain was assessed in 18 tender point localizations consistent with the ACR1990 criteria. To count as a significant tender point, the participant had to indicate pain when a digital-like mechanical pressure was applied. Pain threshold, i.e. when pain was noted, was measured at each of the 18 specified tender points by a valid handheld Somedic algometer (Somedic SenseLab AB, Sweden; the probe had a 1-cm² round rubber application surface, range 0–2000 kPa) [30, 31]. A tender point was deemed positive when the participant perceived pain following a mechanical pressure of 400 kPa or less. The mean of two measurements at each tender point was used for the analysis. The total count of positive tender points was measured for all participants, and the participants needed a tender point count (TPC) >10/18 in order to have fibromyalgia.

Eventually, an algometer score was calculated as the median of the minimum pain-pressure values obtained for each tender point (i.e. median, IQR), but this was not used for classification or diagnostics (see Table 1).

Table 1: Sample characteristics.

		Fibromyalgia	Control	p-Value
Socio-demographic	n	120	62	
Age	mean (SD)	53.1 (10.9)	48.4 (13)	<0.01 ^a
Female (vs. male)	n (%)	119 (99.2)	59 (95.2)	=0.08 ^b
High education (vs. low)	n (%)	35 (29.2)	47 (75.8)	<0.01 ^b
Partner (yes vs. no)	n (%)	88 (73.3) ^c	48 (77.4)	=0.55 ^b
Working/studying (vs. no)	n (%)	34 (28.3) ^c	51 (82.3) ^d	<0.01 ^b
Economy (moderate/good vs. poor)	n (%)	94 (78.3) ^c	57 (91.9)	=0.02 ^b
Clinical characteristics				
Depression (% moderate/serious)	n (%)	20 (17.1) ^c	2 (4.1) ^d	=0.02 ^b
Anxiety (% moderate/serious)	n (%)	18 (15.1) ^c	3 (6.1) ^d	=0.10 ^b
TPC (0–8)	Median (IQR)	14 (11–16) ^c	3 (2–7)	<0.001 ^e
Algometer-score (kPa)	Median (IQR)	323.6 (261.6–371.3) ^c	529.4 (456.5–610.7) ^d	<0.001 ^e
SSS (score 0–12)	Median (IQR)	8 (6–10) ^c	3 (1–5)	<0.001 ^e
WPI (score 0–19)	Median (IQR)	12 (9–15)	5 (1–7)	<0.001 ^e
FS (score 0–31)	Median (IQR)	20 (17–23) ^c	8 (2–12)	<0.001 ^e
Pain (VAS 0–10)	Median (IQR)	7 (5–8)	3 (0–4.8) ^d	<0.001 ^e
Fatigue (VAS 0–10)	Median (IQR)	8 (7–8) ^c	4 (2–5.5) ^d	<0.001 ^e
FIQ-total (0–100) score	Median (IQR)	52.7 (42.3–62.0) ^c	19.2 (6.25–37.3) ^d	<0.001 ^e

^aCalculated using independent samples t-test for parametric variables, mean (standard deviation, SD).

^bn (%) for categorical variables, differences tested with Chi-squares.

^cDo not sum up to 120 in the fibromyalgia group due to missing information (2 for partner, 3 for work/studying, 7 for economy, 3 for depression, 1 for anxiety, 4 for number of tender points, 5 for algometer-score (kPa), 1 for SSS score and FS, 1 for fatigue and 4 missing for FIQ total).

^dDo not sum up to 62 in the control group due to missing information (1 for work/studying; 13 for depression, 13 for anxiety, 1 for algometer-score (kPa), 5 for pain, 5 for fatigue and 10 missing for FIQ total).

^eCalculated using Mann-Whitney test for continuous, not-normally distributed variables, median (interquartile range, IQR).

SSS = symptom severity score; WPI = widespread pain index; FS = fibromyalgia severity (score) = PSD (polysymptomatic distress score) = fibromyalginess. FIQ = fibromyalgia impact questionnaire.

3 Statistics

The data management and the analysis were conducted with the SPSS version 25 (IBMSPSS, Chicago, IL, USA). Descriptive analyses of demographic and clinical characteristics between persons with fibromyalgia and controls were performed with the chi-square statistic for categorical variables and with nonparametric Mann–Whitney U test for continuous variables since the distribution was not normal. *p*-Values below 0.01 were considered statistically significant.

The evaluation of the FSQ included internal consistency reliability for the FS score (sum-score of the entire scale) and separately for SSS and WPI by the coefficient alpha (Cronbach). Test retest intra-reliability analyses of participants answering the FS 1 week apart were performed using intra-class correlation (ICC) (Model Mixed and Type Absolute).

Construct validity included analyses of convergent and discriminant (divergent) validity. To assess

convergent validity the Spearman correlation (*rho*) between the self-report FSQ, tender points (TPs) counting and FIQ was performed. A correlation was strong if *r* was higher than 0.6, moderate if *r* was between 0.3 and 0.6 and low if *r* was below 0.3. To assess discriminant validity the Mann-Whitney U test was used to compare the FS score and subscales of FSQ among those with and without a FM diagnosis.

Receiver Operating Characteristics (ROC) analysis was performed using FS score, and the Area Under the Curve (AUC) was calculated using the ACR1990 as the reference. The sensitivity, the specificity, the positive and negative likelihood ratios (LR+ and LR–) and the accuracy were used to calculate the best cut-off points for the FS score. In addition, the sensitivity, specificity, LR+, LR– and accuracy were established for the 2011 and 2016 FSDC using the ACR1990 criteria as reference. Accuracy is defined as the proportion of the study group correctly classified as positive or negative.

4 Results

One-hundred and twenty-four persons with FM and sixty-four without FM, aged 18–70 years, were invited to the study. Of these, four FM-patients and two controls were excluded due to age >70 and missing data. This left 120 FM patients (119 women) and 62 controls (59 women) to be enrolled in the study, mean age 53.1 (SD 10.9) and 48.4 (SD 13) years, respectively. The socio-demographic and clinical characteristics are shown in Table 1. Age, education and working status revealed statistically significant differences between FM and the control group ($p < 0.01$). For the clinical characteristics, significant differences were found for all variables ($p < 0.001$), except depression and anxiety.

To evaluate the internal consistency of the FSQ, the Cronbach's alfas were measured for FS, WPI and SSS, which revealed 0.904, 0.888 and 0.793, respectively. The Cronbach's alfas for the FM-patients and the controls are shown in Table 2. Test-retest reliability was run on a subsample of the FM group ($n = 27$) performing intraclass correlations (ICC) for FS, WPI, SSS, which showed 0.862, 0.835 and 0.867, respectively. The ICC for the 25 single FS items are shown in Table 3.

Spearman's correlations were calculated in order to test the construct, convergent validity. FSQ (including FS, WPI and SSS) correlated significantly with FIQ total and all its subscales (FIQ function, FIQ overall, FIQ symptoms and the FIQ single items, except item two items [anxiety and depression]). The FS, WPI and SSS correlation coefficients with FIQ total were 0.74, 0.59, 0.85, respectively. TPC also correlated significantly with FIQ total and the major subscales. Results are presented in Table 4.

The percentages of those who met the FSDC 2011 and 2016 criteria in the FM group were significantly greater than in the non-FM group (89.9 vs. 17.7 and 83.2 vs. 14.5, respectively, $p < 0.01$). The sensitivity, specificity, positive and negative likelihood ratio (LR+, LR-) of both the 2011 and 2016 FSDC with ACR1990 FM diagnosis as the

Table 3: Test-retest reliability: intraclass correlation (ICC)^a paired sample correlations for single items and categories of the FS.

	n	ICC
Single items SSS		
Fatigue	27	0.859
Trouble thinking/remembering	27	0.709
Unrefreshing sleep	27	0.773
Pain/cramps in lower abdomen	27	0.494
Depressive symptoms	27	0.879
Headache		0.879
Single items WPI		
Shoulder L	27	0.758
Shoulder R	27	0.830
Hip L	27	0.737
Hip R	27	0.606
Upper arm L	27	0.765
Upper arm R	27	0.757
Lower arm L	27	0.664
Lower arm R	27	0.774
Upper leg L	27	0.845
Upper leg R	27	0.827
Lower leg L	27	0.632
Lower leg R	27	0.560
Neck	27	0.458
Jaw L	27	0.440
Jaw R	27	0.600
Chest	27	0.435
Abdomen	27	0.510
Lower back	27	0.510
Upper back	27	0.695
Scales		
SSS	27	0.867
WPI	27	0.835
FS (SSS + WPI)	27	0.862

^aModel mixed and type absolute.

SSS = symptom severity score; WPI = widespread pain index; FS = fibromyalgia severity (score) = PSD (polysymptomatic distress score) = fibromyalgianess score.

reference were good, comparing FM-patients and controls. The accuracy rate for both criteria was 84%. A receiver operator characteristic (ROC) analysis showed that the optimal cutoff-point for the FS score was 12.5 revealing a 93.9% sensitivity and 68.4% specificity. The AUC was very good (0.860). Results are presented in Table 5.

Table 2: Cronbach's alfa for total sample and by fibromyalgia vs. control.

	Total sample	Fibromyalgia	Control
SSS ^a	0.793	0.784 ^a	0.596
WPI	0.888	0.853	0.796
FS (=SSS + WPI) ^a	0.904	0.774 ^a	0.881

^aMissing information on SSS and FS on 1 participant.

SSS = Symptom Severity Score; WPI = Widespread Pain Index; FS = Fibromyalgia Severity (score) = PSD (polysymptomatic distress score) = "Fibromyalgianess" score.

5 Discussion

We found the Norwegian version of the fibromyalgia survey questionnaire (FSQ-N) to be a valid instrument to diagnose fibromyalgia and applicable in surveys for chronic widespread pain. The 2011/2016 FSDC allow the use of merely

Table 4: Spearman's rho correlations between tender points counting (TPC) and FSQ subgroups versus FIQ total and some major FIQ subclassifications.

	TPC	FIQ-function	FIQ-overall	FIQ-symptoms	FIQ-pain	FIQ-fatigue	FIQ-sleep	FIQ-anxiety	FIQ-depression	FIQ-total
TPC	1	0.51	0.46	0.58	0.58	0.59	0.57	0.20	0.17	0.59
WPI	0.55	0.44	0.52	0.57	0.65*	0.61	0.56	0.27	0.20	0.59
SSS	0.61*	0.64*	0.63*	0.83*	0.73*	0.78*	0.78*	0.47	0.54	0.85*
S-Fatigue	0.51	0.62*	0.54	0.72*	0.67*	0.67*	0.69*	0.29	0.39	0.74*
S-Sleep	0.51	0.59	0.57	0.68*	0.60	0.60	0.79*	0.34	0.36	0.72*
S-Cognition	0.52	0.55	0.53	0.70*	0.64*	0.65*	0.61*	0.38	0.48	0.72*
S-Abdomen	0.43	0.24	0.29	0.37	0.38	0.35	0.35	0.23	0.12	0.36
S-Dep	0.20	0.17	0.30	0.47	0.25	0.22	0.26	0.60	0.73*	0.45
S-Headache	0.41	0.36	0.36	0.40	0.41	0.44	0.41	0.24	0.21	0.42
FS	0.63*	0.55	0.60	0.72*	0.74*	0.73*	0.71*	0.37	0.34	0.74*

A correlation was considered to be strong if r was higher than 0.6*, moderate if r was between 0.3 and 0.6 and low if r was below 0.3.

SSS = symptom severity score; WPI = widespread pain index; FS = fibromyalgia severity (score) = PSD (polysymptomatic distress score) = fibromyalginess; TPC = tender point counts; FS = fibromyalgia severity (score).

Table 5: Receiver Operating Characteristics (ROC) analysis including all participants evaluated by using FS score with the 1990 ACR diagnosis as gold standard, and contingency tables assessment between 2011 FSDC and 2016 FSDC versus ACR1990 criteria ($n = 178$).

	AUC	SS %	SP %	LR+	LR-	ACC %
FS 1990 ACR	0.860					
>9.5		98.0	52.5	2.06	0.04	78
>10.50		98.0	57.5	2.31	0.03	80
>11.50		96.9	62.5	2.58	0.05	81
=12.50		93.9	68.4	2.97	0.09	82
>13.50		90.8	68.7	2.90	0.13	81
>14.50		84.7	72.5	3.08	0.21	79
>15.50		82.7	75.0	3.31	0.23	79
2011 criteria vs. 1990 ACR		93.9	71.3	3.27	0.09	84%
2016 criteria vs. 1990 ACR		88.8	77.5	3.95	0.15	84%

AUC = Areal Under the Curve, SS = sensitivity, SP = specificity, LR+ = positive likelihood, LR- = negative likelihood, ACC = accuracy.

self-reporting, while the ACR2010 also needs a physician confirmation on one item. The intention behind the original ACR2010 and its 2011 revision was to develop simpler tools for clinical diagnosis and research in fibromyalgia without performing tender point examinations, as well as adding a graded severity scale for extra fibromyalgia symptoms [15]. The 2016 FSDC revision highlighted and re-introduced the widespread pain criterion plus the not-exclusive criterion, which states that a “diagnosis of fibromyalgia is valid irrespective of other diagnoses” [4]. It also intended to merge and optimize the clinical diagnostic and research perspectives from the 2010/2011 criteria. Thus, we wanted to examine and validate the Fibromyalgia Survey Questionnaire (FSQ) and its FSDCs for both the 2011 and 2016

versions, in addition to examine the optimal cut-off point of the FS score for the fibromyalgia-diagnosis in Norwegian.

In this study we found the fibromyalgia patients to be a little older than the controls, they had lower education levels and often were working less. The clinical examinations acknowledged differences in tender points counting (TPC), pain thresholds for mechanical pressure (KPa) and fibromyalgia impact (FIQ), as well as WPI, SSS and FS (from the FSQ), indicating more pain and co-morbid symptoms in fibromyalgia than controls, i.e. the same significant contrasts as shown in, e.g. the Persian, Korean, and Spanish validation studies [19, 23, 32]. Our findings indicate that FSQ-N can differentiate between persons with fibromyalgia as defined by the ACR 1990, 2010, 2016 criteria and controls and how to distinguish between FM and non-FM.

The Cronbach's alfa of FS, WPI and SSS, were good (0.904, 0.888 and 0.793, respectively), as well as the test-retest (intra-rater) reliability analysis of the 25 FSQ-N single items, disclosing an acceptable internal consistency of the FSQ-N in line with other validation studies from France, Germany, Iran, Korea and Turkey [22] [17, 19, 21, 23]. The test retest reliability showed an intraclass correlation coefficient (ICC) range from 0.435 to 0.879, i.e. comparable numbers with e.g. Turkish (0.383–0.818) [21] and French (0.600–0.891) [22] numbers for the same items. This supports the assumption that FSQ-N is a reliable questionnaire.

We also performed correlation analyzes to test the construct, convergent validity of FSQ-N (including its FS, WPI and SSS subscales), versus the well-known and valid Fibromyalgia Impact Questionnaire (FIQ) [28]. For these analyses we used Spearman's correlation as a

conservative approach and because some subcategories of FIQ were not normally distributed. FS score correlated significantly and strongly with FIQ total, FIQ symptoms and most FIQ single items (>0.70), except anxiety and depression, while FIQ function and FIQ overall correlated moderately with the FS score. Our FS, WPI and SSS correlation coefficients with FIQ total were 0.74, 0.59 and 0.85, respectively. To compare, the Persian had 0.45, 0.33, 0.45 and the Spanish 0.65, 0.50, 0.66, respectively [19, 32], while the Korean publicized very good WPI and SSS vs FIQ correlations (0.815, 0.854), and the French a single, modest FS versus FIQ correlation ($=0.67$) [22, 23]. The German validation study only ran FS with anxiety and depression correlations [17]. As in the comparable Persian study [19], our FS and SSS scores correlated much better with the FIQ total and symptom scores than with FIQ function, probably because of overlapping sleep, fatigue and pain and depression symptoms, but no function scores.

Our TPC measures correlated somewhat weakly with FS, WPI and SSS compared with the respective Spanish (0.63, 0.55, 0.61 vs. 0.71, 0.69, 0.65) and Korean studies (0.71, 0.69, 0.65) [23, 32] but much better than the Persian (0.11, 0.07, 0.14). In the original FM 2011 study, the TPC versus WPI and SSS correlations were 0.77 and 0.68, respectively [15]. Our TPC correlation with FIQ was 0.59, while, e.g. the Spanish was weaker (Spearman's $r=0.24$), from which they concluded that an increasing number of tender points does not indicate fibromyalgia impact [33]. However, in our study FS score correlated well with both FIQ score and number of TPC, thus supporting the presumption that FSQ-N has good construct, convergent validity for fibromyalgia diagnosis and research according to both fibromyalgia impact and tender points counting.

The number of persons who met the 2011 and 2016 FSDC criteria in the a priori FM group (i.e. those who answered “yes” to the fibromyalgia question in the HUNT3 survey) was significantly greater than in the corresponding a priori non-FM group. These findings are in line with other studies [15, 20, 32].

The ACR1990 criteria were used as a reference versus the two new FM criteria to reveal sensitivity and specificity measures for our a priori FM and non-FM subjects by contingency table analyses. The FSQ-N showed good sensitivity for detecting fibromyalgia using the 2011 and 2016 FSDC (93.9 and 88.8, respectively), as well as moderate specificity (71.3 and 77.5) and positive likelihood ratios (3.27 and 3.95) with acceptable negative likelihood ratios (0.09 and 0.15). These measures showed quite similar sensitivity values as the original 2011 FSDC as well as the Spanish and Turkish versions, and better than, e.g. the Japanese and Persian studies [19, 20]. Nevertheless, a

recent Korean study has probably the best measures until now (sensitivity and specificity with 95% CI, for 2011 and 2016 criteria, respectively): 90.8 (84.5–95.1) and 85.1 (79.2–89.1); 93.1 (87.3–96.7) and 90.7 (85.5–94.0) [23].

Our positive likelihood ratio levels indicated that a positive FSQ-N could occur in fibromyalgia versus non-fibromyalgia with an OR of about 4, which is acceptable, but it is poorer than, e.g. the Spanish results (OR=10.8) [32], while the negative likelihood ratio was quite similar as in our 2011/2016 FSDC findings (0.13 vs. 0.09).

We also measured accuracy, which is defined as “the extent to which a test-value” (e.g. from the FSQ of the 2011/2016 criteria) “reflects or agrees with the reference value” (e.g. ACR1990), i.e. the proportion of correct outcomes of a method [34]. The accuracy scores for both the 2011 and 2016 FSDC vs. ACR1990 were 84% in our contingency table analyses. This is a better outcome than, e.g. the German (72.7%) and Persian studies (75.7), but poorer than the Spanish (89%). Our ROC analyses also indicate that the best cut-off for the FS-scale (WPI + SSS) in classifying fibromyalgia is 12.5 (range 0–31). This replicates the Spanish validation study [32] and confirms the original studies which suggested 13 as an optimal FS cut-off [15, 35], but in contrast to the Persian and Japanese ROC-analyses, with their low 8.5 and 9.5 points, respectively [19, 20]. The “accuracy” term is often used interchangeably with “concordance”, which is not quite correct. While accuracy is measured at a certain cut-off point used for dichotomizing the probability distribution given by logistic model, the “concordance” measures the “overall accuracy” over a range of cutoff points expressed the equivalent term = AUC (Area Under Curve). The AUC in our ROC (receiver operating characteristic) curve was 0.86, i.e. “very good”, but still poorer than the excellent Korean AUC = 0.97 [23] (0.9–1.0 = excellent, 0.8–0.9 = very good, 0.7–0.8 = good). Our AUC ROC curve is shown in Fig. 1.

The 2011/2016 FSDC allow self-reports for both individual diagnostics and epidemiological research. A combined ACR1990 + FSDC approach has been tested, but did not add any benefits to single approaches, revealing only a modest sensitivity (75.6%) [32].

5.1 Limitations and strengths

There were some limitations in the study. The FM patients were little older, had poorer education and were less often working than the controls. This could be explained by the way subjects were recruited, e.g. many controls were employees at outpatient clinics. The male: female ratio was low in both groups. A later study could preferably

Figure 1: Receiver Operating Characteristics (ROC) curve using FS score with the 1990 ACR diagnosis.

have matched the age of the controls at inclusion. The FM patients were not recruited from primary care, but rather from patient organizations. Thus, it is doubtful whether the results of FM severity can be generalized to primary care, and to males. It could be considered a limitation that we did not have a non-painful control group. However, after some discussion we decided not to add it. Zero-pain participants would be very easy to reject as a non-fibromyalgia case without the need of a fibromyalgia examination or questionnaire, so instead we included a control-group with some pain, but without fibromyalgia. Another limitation could be that there is no precise gold standard for diagnosing fibromyalgia, but rather what experts think, and thus that the ACR1990 criteria is not unanimously agreed upon as such. Hence, e.g. Cohen et al. argue that fibromyalgia syndrome is a problem of tautology [36]. Therefore, we preferred to label the ACR1990 criteria a *reference* against the 2016 FSDC instead of a *gold standard*. Eventually, it could be a limitation that we did not use the newer the FIQ revised (FIQ-R), but this is not translated into Norwegian. However, FIQ-R has a good correlation with the original FIQ [37]. Although the 1990 criteria specified digital assessment of tender points, it became recognized with time that the digital examination could be biased and unreliable [38–40]. Specific well-documented algometry examinations were subsequently described and validated [41] to assess

the tender point count. To avoid biased assessment, we used the algometry method in this paper. It is possible that the results would have been somewhat different had we relied on the older 1990 digital method, and it could be a limitation.

The strengths were a fairly large sample size, and an adequate translation procedure according to optimal standards with the original authors participating in the process. The inclusion via the patient organizations could be regarded as a strength, because the researchers then could not interfere with the inclusion and thus selection bias was reduced. The researcher who analyzed the data (ASH) did not see any subjects in the study and took no part in the inclusion or the clinical examination. The patient data were secured with GDPR standards anchored in the NFA patient organization. Another strength was that we evaluated both the 2011 and 2016 FSDC simultaneously with the clinical examination, which also included a history of widespread pain >6 months, pain drawings in electronic body maps and tender points counting (TPC) using an algometer measuring mechanical pressure in kPa for all tender points for both FM and controls. In a sub-sample TPC were assessed by two independent researchers.

5.2 Implications for practice and research

The 2016 FSDC has been validated for both clinical practice and research to provide a quicker, practical approach to identify potential cases than before in Norway, and more practical to use in larger surveys. The online version of this article offers the Norwegian version of FSDC as Supplementary material, see below. Future research could study whether a clinical FM diagnosis would be more valid if a physician or other health-workers monitor and quality-assure the subjects' FSDC self-report scoring process.

6 Conclusion

Our study shows that the Norwegian version of FSQ (FSQ-N) is useful and a valid tool to assess fibromyalgia including its co-morbid core-symptoms such as chronic widespread pain and symptom severity on a continuous scale. The 2016 revision of the 2010/2011 Fibromyalgia Survey Diagnostic Criteria (FSDC) can operate as both a diagnostic instrument for clinical practice and to be applicable for classification in surveys and research in a Norwegian context.

Authors' statements

Research funding: General Practitioner Research Unit, NTNU, Trondheim, Norway; Norwegian Fibromyalgia Association (NFA).

Conflict of interest: All authors and contributors state that there are no conflicts of interest.

Informed consent: All participants signed an informed consent, either electronically or on paper, before inclusion.

Ethical approval: The study was approved by the Regional Committee for Medical and Health Research Ethics (project REK Nord # 2014/938), including fulfillment of the General Data Protection Regulation and in accordance with the Declaration of Helsinki.

References

- [1] Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, Tugwell P, Campbell SM, Abeles M, Clark P, Fam AG, Farber SJ, Fiechtner JJ, Franklin CM, Gatter RA, Hamat D, Lessard J, Lichtbroun AS, Masi AT, McCain GA, et al. The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. *Arthritis Rheum* 1990;33:160–72.
- [2] Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, Russell AS, Russell IJ, Winfield JB, Yunus MB. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care Res (Hoboken)* 2010;62:600–10.
- [3] Clauw DJ. Fibromyalgia: a clinical review. *J Am Med Assoc* 2014;311:1547–55.
- [4] Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Hauser W, Katz RL, Mease PJ, Russell AS, Russell IJ, Walitt B. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin Arthritis Rheum* 2016;46:319–29.
- [5] Fitzcharles MA, Boulos P. Inaccuracy in the diagnosis of fibromyalgia syndrome: analysis of referrals. *Rheumatology (Oxford)* 2003;42:263–7.
- [6] Buskila D, Neumann L, Sibirski D, Shvartzman P. Awareness of diagnostic and clinical features of fibromyalgia among family physicians. *Fam Pract* 1997;14:238–41.
- [7] Crofford LJ, Clauw DJ. Fibromyalgia: where are we a decade after the American College of Rheumatology classification criteria were developed? *Arthritis Rheum* 2002;46:1136–8.
- [8] Yunus M, Masi AT, Calabro JJ, Miller KA, Feigenbaum SL. Primary fibromyalgia (fibrositis): clinical study of 50 patients with matched normal controls. *Semin Arthritis Rheum* 1981;11:151–71.
- [9] Fitzcharles MA, Yunus MB. The clinical concept of fibromyalgia as a changing paradigm in the past 20 years. *Pain Res Treat* 2012;2012:184835.
- [10] Crette S. Fibromyalgia 20 years later: what have we really accomplished? *J Rheumatol* 1995;22:590–4.
- [11] Gunn CC. 'Fibromyalgia'—'what have we created?' (Wolfe 1993). *Pain* 1995;60:349–50.
- [12] Gordon DA. Fibromyalgia—out of control? *J Rheumatol* 1997;24:1247.
- [13] Goldenberg DL. Fibromyalgia syndrome a decade later: what have we learned? *Arch Intern Med* 1999;159:777–85.
- [14] Quintner JL, Cohen ML. Fibromyalgia falls foul of a fallacy. *Lancet* 1999;353:1092–4.
- [15] Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Hauser W, Katz RS, Mease P, Russell AS, Russell IJ, Winfield JB. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR Preliminary Diagnostic Criteria for Fibromyalgia. *J Rheumatol* 2011;38:1113–22.
- [16] Wolfe F, Butler SH, Fitzcharles M, Hauser W, Katz RL, Mease PJ, Rasker JJ, Russell AS, Russell IJ, Walitt B. Revised chronic widespread pain criteria: development from and integration with fibromyalgia criteria. *Scand J Pain* 2019;20:77–86.
- [17] Hauser W, Jung E, Erbsloh-Moller B, Gesmann M, Kuhn-Becker H, Petermann F, Langhorst J, Weiss T, Winkelmann A, Wolfe F. Validation of the Fibromyalgia Survey Questionnaire within a cross-sectional survey. *PLoS One* 2012;7:e37504.
- [18] Carrillo-de-la-Pena MT, Trinanes Y, Gonzalez-Villar A, Romero-Yuste S, Gomez-Perretta C, Arias M, Wolfe F. Convergence between the 1990 and 2010 ACR diagnostic criteria and validation of the Spanish version of the Fibromyalgia Survey Questionnaire (FSQ). *Rheumatol Int* 2015;35:141–51.
- [19] Bidari A, Ghavidel-Parsa B, Amir Maafi A, Montazeri A, Ghalebaghi B, Hassankhani A, Aarabi Y, Haghdoust A. Validation of fibromyalgia survey questionnaire and polysymptomatic distress scale in a Persian population. *Rheumatol Int* 2015;35:2013–9.
- [20] Usui C, Hatta K, Aratani S, Yagishita N, Nishioka K, Kanazawa T, Itoh K, Yamano Y, Nakamura H, Nakajima T, Nishioka K. The Japanese version of the modified ACR preliminary diagnostic criteria for fibromyalgia and the fibromyalgia symptom scale: reliability and validity. *Mod Rheumatol* 2013;23:846–50.
- [21] Yanmaz MN, Atar S, Bicer M. The reliability and validity of the Turkish version of fibromyalgia survey diagnostic criteria and symptom severity scale. *J Back Musculoskelet Rehabil* 2016;29:287–93.
- [22] Fitzcharles MA, Ste-Marie PA, Panopalis P, Menard H, Shir Y, Wolfe F. The 2010 American college of rheumatology fibromyalgia survey diagnostic criteria and symptom severity scale is a valid and reliable tool in a French speaking fibromyalgia cohort. *BMC Musculoskelet Disord* 2012;13:179.
- [23] Kang JH, An M, Choi SE, Xu H, Park DJ, Lee JK, Lee SS. Performance of the revised 2016 fibromyalgia diagnostic criteria in Korean patients with fibromyalgia. *Int J Rheum Dis* 2019;22:1734–40.
- [24] The SAGE Encyclopedia of Communication Research Methods. 2017. doi:10.4135/9781483381411.
- [25] Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, Bratberg G, Heggland J, Holmen J. Cohort profile: the HUNT study, Norway. *Int J Epidemiol* 2013;42:968–77.
- [26] Reme SE, Lie SA, Eriksen HR. Are 2 questions enough to screen for depression and anxiety in patients with chronic low back pain? *Spine* 2014;39:E455–62.
- [27] Burckhardt CS, Clark SR, Bennett RM. The fibromyalgia impact questionnaire – development and validation. *J Rheumatol* 1991;18:728–33.
- [28] Bennett R. The fibromyalgia impact questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. *Clin Exp Rheumatol* 2005;23(5 Suppl 39):S154–62.

- [29] Jaatun EA, Haugen DF, Dahl Y, Kofod-Petersen A. Designing a reliable pain drawing tool: avoiding interaction flaws by better tailoring to patients' impairments. *Pers Ubiquitous Comput* 2015;19:635–48.
- [30] Chesterton LS, Sim J, Wright CC, Foster NE. Interrater reliability of algometry in measuring pressure pain thresholds in healthy humans, using multiple raters. *Clin J Pain* 2007;23:760–6.
- [31] Bernhardt O, Schiffman EL, Look JO. Reliability and validity of a new fingertip-shaped pressure algometer for assessing pressure pain thresholds in the temporomandibular joint and masticatory muscles. *J Orofac Pain* 2007;21:29–38.
- [32] Segura-Jimenez V, Aparicio VA, Alvarez-Gallardo IC, Soriano-Maldonado A, Estevez-Lopez F, Delgado-Fernandez M, Carbonell-Baeza A. Validation of the modified 2010 American College of Rheumatology diagnostic criteria for fibromyalgia in a Spanish population. *Rheumatology (Oxford)* 2014;53:1803–11.
- [33] Henriksen M, Lund H, Christensen R, Jespersen A, Dreyer L, Bennett RM, Danneskiold-Samsøe B, Bliddal H. Relationships between the fibromyalgia impact questionnaire, tender point count, and muscle strength in female patients with fibromyalgia: a cohort study. *Arthritis Rheum* 2009;61:732–9.
- [34] Medical Dictionary for the Health Professions and Nursing. 2012 Available from: <https://medical-dictionary.thefreedictionary.com/accuracy>.
- [35] Ferrari R, Russell AS. A questionnaire using the modified 2010 American College of Rheumatology criteria for fibromyalgia: specificity and sensitivity in clinical practice. *J Rheumatol* 2013;40:1590–5.
- [36] Cohen ML, Quintner JL. Fibromyalgia syndrome, a problem of tautology. *Lancet* 1993;342:906–9.
- [37] Bennett RM, Friend R, Jones KD, Ward R, Han BK, Ross RL. The revised fibromyalgia impact questionnaire (FIQR): validation and psychometric properties. *Arthritis Res Ther* 2009;11:R120.
- [38] Wilson HD, Starz TW, Robinson JP, Turk DC. Heterogeneity within the fibromyalgia population: theoretical implications of variable tender point severity ratings. *J Rheumatol* 2009;36:2795–801.
- [39] Gracely RH, Grant MA, Giesecke T. Evoked pain measures in fibromyalgia. *Best Pract Res Clin Rheumatol* 2003;17:593–609.
- [40] Petzke F, Gracely RH, Park KM, Ambrose K, Clauw DJ. What do tender points measure? Influence of distress on 4 measures of tenderness. *J Rheumatol* 2003;30:567–74.
- [41] Okifuji A, Turk DC, Sinclair JD, Starz TW, Marcus DA. A standardized manual tender point survey. I. Development and determination of a threshold point for the identification of positive tender points in fibromyalgia syndrome. *J Rheumatol* 1997;24:377–83.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/sjpain-2020-0002>).