RHEUMATOLOGY

Rheumatology 2014;53:1803–1811
doi:10.1093/rheumatology/keu169
Advance Access publication 14 May 2014

Original article

Validation of the modified 2010 American College of Rheumatology diagnostic criteria for fibromyalgia in a Spanish population

Víctor Segura-Jiménez[1], Virginia A. Aparicio[1,2], Inmaculada C. Álvarez-Gallardo[1], Alberto Soriano-Maldonado[1], Fernando Estévez-López[1], Manuel Delgado-Fernández[1] and Ana Carbonell-Baeza[3]

Abstract

Objective. The aim of this study was to validate the modified 2010 ACR preliminary criteria for FM in a Spanish population.

Methods. Five hundred and seventy-nine (550 women) FM and 294 (240 women) control participants were enrolled in the study. FM patients were previously diagnosed by a rheumatologist. All participants underwent both the 1990 ACR criteria (1990c) and the modified 2010 ACR criteria (m-2010c).

Results. The tender points count showed correlations of 0.69, 0.65 and 0.71 with the widespread pain index (WPI), symptoms severity (SS) and polysymptomatic distress (PSD) scales, respectively (all $P < 0.001$). The WPI, SS and PSD showed greater correlations with impact of FM health-related quality of life, general fatigue and depression than the tender points count. The 1990c showed sensitivity and specificity values of 84.1 and 97.6, respectively, whereas the m-2010c showed values of 88.3 and 91.8, respectively. Both criteria showed the same overall accuracy, with a value of 0.89. When the 1990c and m-2010c were combined and patients had to satisfy one of two criteria to be diagnosed with FM, the sensitivity, specificity and accuracy of questionnaires were 96.7, 89.8 and 0.94, respectively. The original cut-off points (WPI $\geqslant$ 7, SS $\geqslant$ 5 and PSD $\geqslant$ 12) showed the best test characteristics in the present study.

Conclusion. The m-2010c, with the same cut-off points as the original version, are a valid tool for the diagnosis of FM in our population. Whenever possible, the combination of the 1990c and m-2010c is recommended (patients have to meet one of the two criteria to be diagnosed), since this approach showed the best diagnostic characteristics.

Key words: tender points, questionnaire, polysymptomatic distress scale, widespread pain index, symptom severity, quality of life, impact of fibromyalgia, fatigue, sensitivity, specificity.

Introduction

FM has become a worrisome health condition in our modern society. This condition was first referred to as fibrositis and was mainly centred on diffuse pain [1]. A few years later, Yunus *et al.* [2] called it fibromyalgia and proposed a set of criteria for its diagnosis, including tender points and the presence of different symptoms. In 1990 the ACR reported the first criteria to differentiate FM from other chronic widespread pain syndromes [3]. Twenty years later, new, presumably improved ACR preliminary FM criteria have been released [4].

The 1990 ACR criteria (hereinafter referred to as 1990c) required widespread pain for at least 3 months and the presence of 11 of 18 tender points [3]. As time passed, objections to the 1990c developed, on the grounds that the presence of different tender points cannot be an objective assessment of whole body pain [5]. Furthermore,

[1]Department of Physical Activity and Sport, Faculty of Sport Sciences, University of Granada, [2]Department of Physiology, Faculty of Pharmacy, University of Granada, Granada and [3]Department of Physical Education, Faculty of Sciences Education, University of Cádiz, Cádiz, Spain.

Submitted 21 October 2013; revised version accepted 5 March 2014.

Correspondence to: Víctor Segura-Jiménez, Department of Physical Activity and Sport, Faculty of Sport Science, University of Granada, Carretera de Alfacar, s/n, 18011 Granada, Spain.
E-mail: vsegura@ugr.es

digital palpation instead of algometry is the most widely used method among examiners [6]. Without an objective instrument, the application of an equal pressure of 4 kg is therefore doubtful [5, 6]. Finally, FM is defined as a complex multidimensional pain disorder [4, 7] with the inclusion of equally presumably important non-pain symptoms, such as fatigue, stiffness, depression and cognitive problems, among other complaints [7, 8]. The range of these symptoms is therefore another important weakness of the 1990c [4, 5].

In 2010 the ACR released new diagnostic criteria that simplified clinical diagnosis by avoiding the requirement of examination of tender points [4]. With the new 2010 ACR criteria, FM has turned into a systemic symptom-based condition rather than the previous peripheral pain-defined condition [5]. Because most of the 2010 ACR items were obtained by self-administration, the criteria were modified in 2011 to allow complete self-administration [9]. These new criteria, which we shall call the m-2010c, eliminated the physicians' subjective assessments, thus making it a self-administered questionnaire suitable for epidemiological studies [9].

The m-2010c represent an alternative FM assessment method. However, it is not known whether the m-2010c are a valid diagnostic tool in different populations, and further investigations in different countries have been requested. To date, the 2010 ACR criteria have been validated in English- [4], Japanese- [10], French- [11] and Iranian- [12] speaking populations, whereas the m-2010c have been validated in English- [9] and Japanese- [13] speaking populations. To the best of our knowledge, no previous research has studied the validity of the m-2010c in Spain. Therefore we aimed to validate the m-2010c for FM in a representative population from southern Spain.

Patients and methods

Participants

FM patients were recruited from various FM associations via e-mail, letter or telephone. We also asked those FM patients interested in participating to recruit a healthy individual (control) of similar age, socio-demographic characteristics and demographic area in order to carry out appropriate comparisons between groups. We additionally contacted control participants via e-mail and Internet advertisements. All interested participants ($n = 960$) signed a written informed consent after receiving detailed information about the aims and study procedures. The study assessments were carried out between November 2011 and January 2013. The inclusion criteria for FM participants were (i) previous diagnosis of FM by a rheumatologist (patients were asked to provide their medical records to confirm their previous diagnosis) and (ii) no acute or terminal illness (such as cancer, stroke, recent cardiopathy, severe coronary disease, schizophrenia or any other disabling injury) or severe dementia [mini mental state examination (MMSE) < 10] [14]. The inclusion criteria for control participants were (i) no previous diagnosis of FM by a rheumatologist and (ii) no acute or terminal illness or severe dementia (MMSE < 10). A total of 39 participants were excluded from the study. One had an MMSE score <10 and the other 38 participants suffered from pain, although they had not previously visited a rheumatologist for a FM diagnosis. A total of 921 participants from southern Spain were enrolled in the study, which was reviewed and approved by the Ethics Committee of the Hospital Virgen de las Nieves, Granada, Spain.

Outcome measures

We assessed 18 tender points according to the 1990c [3] using a standard pressure algometer (FPK 20; Wagner Instruments, Greenwich, CT, USA). The mean of two measurements at each tender point was used for the analysis. A tender point scored as positive when the patient noted pain at a pressure $\leqslant 4\,kg/cm^2$. The total count of positive tender points (tender points count) was recorded for each participant. An algometer score was calculated as the sum of the minimum pain-pressure values obtained for each tender point.

The questionnaire for the m-2010c [9] is composed of two scales. The widespread pain questionnaire asked participants to grade whether (or not) they had pain or tenderness over the previous week in 19 body areas (shoulder girdle, hip, jaw, upper arm, upper leg, lower arm and lower leg, on the right and the left side of the body, separately, and additionally neck, chest, abdomen, upper back and lower back). Each item was scored as 0 or 1. The minimum total score of the widespread pain index (WPI) was 0 and the maximum total score was 19. The symptom scale questionnaire asked participants to indicate the severity of fatigue, trouble thinking or remembering and waking up tired (unrefreshed) over the previous week. The possible values were 0 (no problem), 1 (slight or mild problems, generally mild or intermittent), 2 (moderate, considerable problems, often present and/or at a moderate level) and 3 (severe, continuous, life-disturbing problems). Patients were also asked to answer whether (or not) they had had pain or cramps in the lower abdomen, depression or headache during the previous 6 months. The minimum total score of symptom severity (SS) was 0 and the maximum total score was 12. The WPI and SS were subsequently summed into a 0–31 index originally called the fibromyalgianess scale [9] and subsequently termed the polysymptomatic distress (PSD) scale [7]. The original 2010 ACR criteria study and the corresponding questionnaire have been previously translated to Spanish (see http://www.institutferran.org/documentos/WPI+SS-PACIENTES.pdf and http://www.institutferran.org/documentos/2010_ACR_FM_TRAD_FINAL.pdf). We used this questionnaire, with the exception that we adapted the second part of the SS to the m-2010c: the physicians' estimation of the SS score was eliminated and replaced with three dichotomous yes/no responses regarding the presence of pain or cramps in the lower abdomen, depression or headache during the previous 6 months, as explained above. The questionnaire was self-administered and patients obtained directions

from the researchers when they did not understand the questionnaire instructions.

The revised Fibromyalgia Impact Questionnaire (FIQR) is a self-administered questionnaire comprising 21 individual questions with a rating scale of 0–10. The questions compose three different domains: function, overall impact and symptoms score (range 0–30, 0–20 and 0–50, respectively) [15, 16]. The FIQR total score ranges from 0 to 100, with a higher score indicating a greater impact of the condition on the person's life.

The 36-item Short Form Health Survey (SF-36) is a generic instrument for assessing health-related quality of life [17, 18]. Its 36 items are grouped into eight dimensions: physical functioning, physical role, body pain, general health, vitality, social functioning, emotional role and mental health. The scores range from 0 to 100 for every dimension, and higher scores indicate better health.

The Multidimensional Fatigue Inventory (MFI) measures fatigue severity. It comprises five subscales: general fatigue, physical fatigue, mental fatigue, reduced activity and reduced motivation [19, 20]. Each subscale includes four items on a 5-point Likert scale. Scores on each subscale range from 4 to 20, with higher scores indicating greater fatigue. In the present study we only focused on general fatigue, which includes general statements about fatigue and decreased functioning.

The Beck Depression Inventory II (BDI-II) was used to assess depression severity [21, 22]. It contains 21 items and the score ranges from 0 to 63, with a higher score indicating greater depression.

The MMSE was used to assess cognitive capacity and the severity of dementia for the exclusion criteria [14, 23]. Five areas of cognitive functioning were assessed: orientation, immediate memory, attention/concentration, delayed recall and language. The score ranges from 0 to 30, with a lower score indicating a greater state of dementia.

Statistical analysis

Differences in socio-demographic variables between groups were calculated with analysis of variance (ANOVA) or the chi-square test when appropriate. Differences in clinical variables were calculated with analysis of covariance (ANCOVA), adjusting for all significant socio-demographic variables. The relationship between the tender points count, the new criteria scales and the other important study variables was studied using Pearson's correlation coefficient (r_p). Cronbach's α was used to measure the internal consistency for the m-2010c. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR) and accuracy for the 1990c, the m-2010c and the combination of both criteria were calculated. The kappa and delta indexes [24] were additionally used to measure the agreement between the gold standard (clinical/rheumatologist diagnosis) and FM criteria. New cut-off points that could represent better test characteristics than those presented in the original m-2010c were studied by means of receiving operator characteristic (ROC) analysis. Significance was set at $P < 0.05$. The Statistical Package for Social Sciences, version 20.0 (SPSS Statistics for Windows, IBM, Armonk, NY, USA) was used.

Results

A total of four participants did not assist to the tender points assessment and 44 participants did not fill out the questionnaires. The final study sample comprised 873 participants. According to the clinical (rheumatologist) diagnosis, 579 (550 women) participants became part of the FM group, whereas 294 (240 women) participants were included in the control group.

The socio-demographic characteristics of the study groups are shown in Table 1. Age, sex, educational status and occupational status were statistically different between the FM and control groups (all $P < 0.001$). The clinical characteristics of the study groups are shown in Table 2. A significant difference in all the clinical variables between the FM and control groups was observed (all $P < 0.001$), except in the MMSE ($P > 0.05$).

To further characterize the relationship between the tender points count, the new criteria scales and the study variables, Pearson correlations are shown in Table 3. The correlations between the tender points count and the WPI, SS and PSD were 0.69, 0.65 and 0.71, respectively (all $P < 0.001$). The scales of the new criteria (WPI, SS and PSD) showed greater correlations with all the study variables than the tender points count.

We applied the 1990c and the m-2010c to participants enrolled in the study. By clinical diagnosis at entry, 63.6% of the total participants were diagnosed with FM. Using 1990c, the percentage of total participants with FM was 58.4, whereas with m-2010c it was 61.0 (Table 4). Among participants completing the FM clinical diagnosis, 84.1% satisfied the 1990c and 88.3% satisfied the m-2010c. Among control participants, 2.4% were diagnosed with FM according to the 1990c and 8.2% according to the m-2010c.

The all-item internal consistency of the m-2010c showed a Cronbach's α of 0.93. The sensitivity, specificity, PPV, NPV, PLR, NLR and accuracy of the 1990c and m-2010c were calculated using the clinical diagnosis as the gold standard (Table 5). The m-2010c showed a greater sensitivity but lower specificity than the 1990c (88.3 *vs* 84.1 and 91.8 *vs* 97.6, respectively). Both criteria showed the same overall accuracy, with a value of 0.89. We further studied the combination of both the 1990c and m-2010c. When patients had to satisfy both criteria to be diagnosed (1990c + m-2010c), the sensitivity was low (75.6), despite the fact that the specificity was almost perfect (99.7). When patients had to satisfy one of the criteria to be diagnosed (1990c or m-2010c), the sensitivity and specificity were both very high (96.7 and 89.8, respectively) and the accuracy, with a value of 0.94, was higher than all the previous options. The kappa and delta values for the 1990c and m-2010c were similar and overall they were good (>0.75). The 1990c + m-2010c combination showed the lowest kappa value, whereas the 1990c or

Table 1 Socio-demographic characteristics of the study groups

	FM ($n = 579$)		Control ($n = 294$)		*P*-value
Age, mean (S.D.), years	51.9	(8.1)	49.1	(10.4)	<0.001
Sex					
Men	29	(5.0)	54	(18.4)	<0.001
Women	550	(95.0)	240	(81.6)	
Marital status					
Married	428	(73.9)	209	(71.3)	0.080
Single	58	(10.0)	46	(15.7)	
Separated	22	(3.8)	13	(4.4)	
Divorced	44	(7.6)	14	(4.8)	
Widow	27	(4.7)	11	(3.8)	
Educational status					
No studies	62	(10.7)	16	(5.4)	<0.001
Primary school	283	(48.9)	109	(37.1)	
Professional training	86	(14.9)	53	(18.0)	
Secondary school	67	(11.6)	50	(17.0)	
University medium degree	52	(9.0)	32	(10.9)	
University higher degree	29	(5.0)	34	(11.6)	
Current occupational status					
Working full time	89	(15.4)	92	(31.3)	<0.001
Working part time	61	(10.5)	36	(12.2)	
Housewife	174	(30.1)	79	(26.9)	
Student	5	(0.9)	7	(2.4)	
Retired/pensioner	23	(4.0)	16	(5.4)	
Retired/incapacity pension	79	(13.6)	13	(4.4)	
Sick leave	44	(7.6)	2	(0.7)	
Unemployed	104	(18.0)	49	(16.7)	

Values are *n* (%) unless otherwise indicated. Differences in socio-demographic variables tested by analysis of variance (ANOVA) or chi-squared test when appropriate.

Table 2 Clinical characteristics of the study groups

	FM ($n = 579$)		Control ($n = 294$)		*P*-value
	Mean	(95% CI)	Mean	(95% CI)	
Tender points count (0–18)	15	(14.4, 15.1)	3	(2.6, 3.7)	<0.001
Algometer score (18–144)	51	(49.5, 53.2)	108	(105.1, 110.5)	<0.001
WPI total (0–19)	13	(13.2, 13.8)	4	(3.6, 4.5)	<0.001
SS score total (0–12)	8	(7.7, 8.0)	2	(2.1, 2.7)	<0.001
PSD scale (0–31)	21	(20.9, 21.8)	6	(5.8, 7.1)	<0.001
FIQR total score (0–100)	64	(62.1, 65.0)	24	(15.6, 32.4)	<0.001
SF-36					
Physical function (0–100)	40	(38.8, 42.1)	79	(77.0, 81.6)	<0.001
Physical role (0–100)	35	(33.2, 36.9)	79	(76.0, 81.3)	<0.001
Bodily pain (0–100)	23	(21.1, 24.3)	65	(62.3, 66.8)	<0.001
General health (0–100)	30	(28.3, 31.2)	63	(61.1, 65.2)	<0.001
Vitality (0–100)	24	(22.6, 25.8)	62	(59.5, 64.2)	<0.001
Social functioning (0–100)	45	(42.8, 46.9)	81	(78.3, 84.2)	<0.001
Emotional role (0–100)	57	(54.4, 58.7)	84	(80.6, 86.9)	<0.001
Mental health (0–100)	46	(44.6, 48.1)	70	(67.3, 72.2)	<0.001
MFI general fatigue (4–20)	18	(17.5, 18.1)	10	(9.9, 10.7)	<0.001
BDI-II (0–63)	26	(25.1, 26.9)	10	(9.1, 11.7)	<0.001
MMSE (0–30)	28	(27.7, 28.0)	28	(27.9, 28.4)	0.057

Differences in clinical variables tested by analysis of covariance (ANCOVA), adjusting for all significant socio-demographic variables. WPI: widespread pain index; SS: symptom severity; PSD: polysymptomatic distress; FIQR: FM impact questionnaire; SF-36: 36-item Short Form Health Survey; MFI: multidimensional fatigue inventory; BDI-II: Beck depression inventory II; MMSE: mini mental state examination.

Table 3 Pearson's correlations between key study variables for all participants ($n = 871$)

	Tender points count	WPI total	SS score total	PSD scale
Tender points count	1.00	0.69	0.65	0.71
WPI total	0.69	1.00	0.77	0.97
SS score total	0.65	0.77	1.00	0.90
PSD scale	0.71	0.97	0.90	1.00
FIQR total score	0.24	0.50	0.66	0.65
SF-36 physical function	−0.61	−0.69	−0.71	−0.74
SF-36 physical role	−0.61	−0.72	−0.76	−0.78
SF-36 bodily pain	−0.66	−0.76	−0.78	−0.81
SF-36 general health	−0.61	−0.69	−0.73	−0.75
SF-36 vitality	−0.62	−0.69	−0.74	−0.75
SF-36 social functioning	−0.52	−0.61	−0.73	−0.69
SF-36 emotional role	−0.40	−0.53	−0.61	−0.59
SF-36 mental health	−0.44	−0.53	−0.61	−0.59
MFI general fatigue	0.64	0.70	0.76	0.77
BDI-II	0.51	0.60	0.70	0.67

WPI: widespread pain index; SS: symptom severity; PSD: polysymptomatic distress; FIQR: revised FM impact questionnaire; SF-36: 36-item Short Form Health Survey; MFI: multidimensional fatigue inventory; BDI-II: Beck Depression Inventory II. All *P*-values <0.001.

Table 4 FM prevalence according to clinical diagnosis and 1990, 2010 and modified 2010 ACR criteria

	FM by clinical diagnosis	FM by 1990 ACR criteria	FM by 2010 or modified 2010 ACR criteria
Spanish modified 2010 ACR criteria			
All participants ($n = 910$)	**63.6**	58.4	61.0
FM ($n = 579$)	**100**	84.1	88.3
Control ($n = 294$)	**0**	2.4	8.2
Original modified 2010 ACR criteria [9]			
All participants ($n = 7233$)	**10.1**	—	25.4
FM ($n = 729$)	**100**	—	60.0
Control ($n = 6504$)	**0**	—	21.6
Japanese modified 2010 ACR criteria [13]			
All participants ($n = 693$)	—	**66.7**	44.0
FM ($n = 462$)	—	**100**	64.0
Control ($n = 231$)	—	**0**	4.0
Original 2010 ACR criteria [4]			
All participants ($n = 514$)	**50.2**	—	38.1
FM ($n = 258$)	**100**	—	74.5
Control ($n = 256$)	**0**	—	2.0
Japanese 2010 ACR criteria [10]			
All participants ($n = 137$)	—	**68.6**	59.1
FM ($n = 94$)	—	**100**	82.0
Control ($n = 43$)	—	**0**	9.0
Iranian 2010 ACR criteria [12]			
All participants ($n = 278$)	**60.4**	47.8	38.4
FM ($n = 168$)	**100**	71.4	58.9
Control ($n = 110$)	**0**	11.7	7.2

The gold standard selected in each validation study is highlighted in bold.

m-2010c showed the best kappa and delta values from all options (see Table 5).

We analysed different cut-off points for the WPI and the SS score. We did not observed new cut-off points with better characteristics with maximal area under the curve than those provided in the original m-2010c (WPI $\geqslant 7$ and SS $\geqslant 5$ or WPI 3–6 and SS $\geqslant 9$) (data not shown). Comparisons between the FM group and the control

Table 5 Test characteristics of ACR criteria for classifying FM using clinical diagnosis as the gold standard ($n = 873$)

Criteria	Sensitivity	Specificity	PPV	NPV	PLR	NLR	Accuracy	Kappa	Delta
1990 ACR criteria	84.1	97.6	98.6	75.7	35.3	0.16	0.89	0.76	0.83
m-2010 ACR criteria	88.3	91.8	95.5	79.9	10.8	0.13	0.89	0.77	0.80
1990 or m-2010 ACR criteria	96.7	89.8	94.9	93.3	9.5	0.04	0.94	0.87	0.89
1990 + m-2010 ACR criteria	75.6	99.7	99.8	67.5	222.4	0.24	0.84	0.67	0.81

PPV: positive predictive value; NPV: negative predictive value; PLR: positive likelihood ratio; NLR: negative likelihood ratio; m-2010: modified 2010 ACR criteria.

Table 6 Test characteristics of the PSD scale based on receiver operator characteristics analysis ($n = 873$)

PSD	Sensitivity	Specificity	PPV	NPV	PLR	NLR	Accuracy
10	97.9	78.2	89.9	95.0	4.5	0.03	0.91
11	97.1	82.3	91.5	93.4	5.5	0.04	0.92
12	95.7	84.7	92.5	90.9	6.3	0.05	0.92
13	93.6	87.4	93.6	87.4	7.4	0.07	0.92
14	91.7	89.1	94.3	84.5	8.4	0.09	0.91
15	90.2	91.8	95.6	82.6	11.0	0.11	0.91
16	86.7	92.9	96.0	78.0	12.1	0.14	0.89

PSD: polysymptomatic distress; PPV: positive predictive value; NPV: negative predictive value; PLR: positive likelihood ratio; NLR: negative likelihood ratio.

group were carried out using ROC analyses for the PSD. Several cut-off scores for the PSD, together with the sensitivity, specificity, PPV, NPV, PLR, NLR and accuracy are shown in Table 6. The best cut-off scores for the PSD were 12 and 13, with a sensitivity of 95.7 and 93.6, respectively, a specificity of 84.7 and 87.4, respectively, and an accuracy of 0.92. A graphical representation of the results from the ROC analyses is presented in supplementary Fig. S1, available at *Rheumatology* Online.

All analyses were repeated with age-matched groups (35–65 years) and sex-separated groups, and the results remained unchanged.

Discussion

The present study showed acceptable validity of the m-2010c as a FM diagnostic tool. When clinical diagnosis was applied as the gold standard, the m-2010c showed greater sensitivity but lower specificity than the 1990c. The cut-off score (WPI $\geqslant$ 7 and SS $\geqslant$ 5 or WPI 3–6 and SS $\geqslant$ 9) published in the original version perfectly fitted our sample and no different cut-off scores showed a substantial improvement. The PSD best cut-off points were 12 and 13, similar to the original version [4, 9]. The combination of both 1990c and m-2010c showed the best test characteristics for the diagnosis of FM.

Age, sex, educational status and occupational status differed between the FM and control groups. Lower levels of education have been associated with an increased prevalence of pain [25] and a higher prevalence rate of FM [26]. Furthermore, FM directly impacts work ability, which implies an indirect economic repercussion. Lower incomes have also been related to a higher prevalence of FM [25–27]. As expected, clinical variables were statistically different between the FM and control groups. Furthermore, the differences observed in tender points count, SS, WPI and PSD between the FM and control groups speaks favourably about the internal consistency of both the 1990c and the m-2010c.

Higher correlations were observed between the m-2010c and various study variables (FIQR, SF-36, MFI and BDI-II) than those obtained for the 1990c and the aforementioned study variables. In line with these results, it has been previously stated that the m-2010c and FIQR can be somewhat compared [9]. We showed a low correlation ($r_p = 0.24$) between the tender points count and FIQR, which indicates that the tender points count by itself may be a poor indicator of FM impact. These results have been shown previously with the FIQ [28, 29]. However, the correlation between the PSD and FIQR was stronger ($r_p = 0.65$) than that observed between the tender points count and the FIQR. This fact, together with the moderate to high correlation with general fatigue and depression (two other highly important FM symptoms [8]) emphasizes the concept of FM as a multisymptom dimensional condition.

There is no gold standard for the diagnosis of FM, which represents a problem when assessing the validity of new criteria. In the absence of this gold standard, an expert consensus was first adopted by Wolfe *et al*. [3] in the 1990c validation study. Subsequently this approach was used in both the 2010 original version and m-2010c [4, 9]. Following this approach, we adopted clinical diagnosis as the gold standard to study the validity of the m-2010c, as previously done in a recent FM criteria validation study in an Iranian population [12].

From the total number of FM participants recruited and diagnosed by rheumatologists, the m-2010c positively identified 88.3%, whereas the 1990 criteria positively identified 84.1%. However, the m-2010c incorrectly classified a higher percentage of control participants as having FM than the 1990 criteria (8.2 *vs* 2.4%). Compared with previous validation studies of the 2010 ACR criteria [4, 10, 12] and m-2010c [9, 13], our results were most accurate when positively identifying FM patients, whereas the incorrect classification of control participants was similar to that presented in previous studies. We further studied the possibility of using both the 1990c and the m-2010c together. Two options were available: (i) participants had to satisfy both criteria to be diagnosed as having FM or (ii) participants had to meet one of two criteria. Although the approach explained in the first option is excellent to reject those without FM (specificity = 99.7), the ability to detect patients with FM was fairly low (sensitivity = 75.6) compared with the 1990c and m-2010c by themselves, which is not acceptable. However, the second option revealed some surprising results, with a sensitivity of 96.7, specificity of 89.9 and accuracy similar for all options. The kappa and delta values for both the 1990c and m-2010c indicated good agreement between these criteria and the clinical diagnosis. Following the same pattern as above, the 1990c or m-2010c combination showed the best agreement. This combination of criteria had not previously been reported, but it seems that when the combination of the two criteria is available, satisfying one of two criteria shows the best overall characteristics to meet the diagnosis of FM. As recently shown, FM patients might not necessarily fulfil the tender points criteria to be diagnosed [30], which agrees with our findings. Furthermore, this criteria combination might help to identify more homogeneous subgroups of patients (e.g. those fulfilling only 1990c, m-2010c, or both criteria). In practice, this would imply retaining clinical and objective review of tender points examination, which would not be ideal in relation to certain types of study such as surveys and large epidemiological studies. However, this would be reasonable (and more accurate overall according to the results of the present study) in the context of clinical practice. Perhaps the m-2010c might be used a priori, due to its simplicity, quickness and self-administration; and for those patients not fulfilling these criteria, the 1990c could provide additional information in order to reach a final diagnosis.

The understanding of FM as a multidimensional disorder raises the importance of the PSD as an FM scale [7]. FM patients are a heterogeneous group and thus their symptomatology may vary between different populations, regions and/or countries [7]. Therefore we studied the possibility of new cut-off points that were able to improve the diagnostic accuracy of the m-2010c. We did not observe any improvement with different cut-off points than those proposed by Wolfe *et al.* [4] in the original version. The PSD has been identified as an important index that might allow FM to be mapped on a dimensional or continuum scale [7, 9]. We further investigated the sensitivity, specificity, PPV, NPV, PLR, NLR and accuracy of different cut-off points than those suggested in the original version [9]. The cut-off points of 12 and 13 were those that best fitted the present study. Both cut-offs showed high sensitivity, specificity and almost identical accuracy. Choosing the cut-off score of 12, the sensitivity increases by 1.9% compared with a cut-off score of 13 and the specificity only decreases by 2.7%. Moreover, this is the original cut-off score proposed for the PSD in the original m-2010c validation [9], and it perfectly fitted our population sample. Previous studies have shown an increase in accuracy with the m-2010c when the cut-off score is changed, reflecting the differences between patients with FM from different countries [10, 12, 13]. However, we found that the original cut-off score perfectly fitted our population and the m-2010c can be used in its original form.

A limitation of the current study is that we did not have an expert consensus as in Wolfe *et al.* [3] and therefore we chose the clinical diagnosis as the gold standard. It is possible that some clinical diagnoses were erroneous, since the clinical diagnosis is not a real gold standard. Although the 2010 ACR criteria have been scientifically translated to Spanish, there is not a Spanish adaptation and psychometric properties study of the m-2010c. The male sample size was low compared with the women's sample size; nevertheless, it is consistent with the general sex prevalence of FM. The study groups showed age differences and women and men were analysed together; however, we repeated all the analyses with age-matched and sex-separated groups, obtaining the same results showed in the present study. The main strength of the present study is its large sample size, which was from southern Spanish population with FM. We also used both the 1990c and the m-2010c in our study sample, which allowed us to compare both criteria and study the usefulness of criteria combinations.

Conclusions

The present study showed the validity of the m-2010c as a FM diagnostic tool in Spain. We suggest using the same cut-off points as the original version. However, the results suggest that using both the 1990c and the m-2010c, and meeting one of the two criteria, might be a better option than using the m-2010c alone. Future studies should examine this combination of criteria.

Rheumatology key messages

- The modified 2010 ACR preliminary criteria are valid for use to diagnose FM in the Spanish population.
- For FM diagnosis, we recommend use of the cut-off points suggested in the modified 2010 ACR preliminary criteria.
- Combination of 1990 and modified 2010 ACR criteria showed the best characteristics for FM diagnosis.

Acknowledgements

We thank all the members from the CTS-545 research group involved in the field work. We also gratefully acknowledge all the study participants for their collaboration.

Funding: This study was supported by the Spanish Ministry of Science and Innovation (I+D+I DEP2010-15639; grant BES-2011-047133); the Consejeria de Turismo, Comercio y Deporte (CTCD-201000019242-TRA); the Spanish Ministry of Education (AP-2010-0963, FPU12/00963); Granada Research of Excelence Initiative on Biohealth (GREIB), Campus BioTic, University of Granada, Spain; and the European University of Madrid, Escuela de Estudios Universitarios Real Madrid (2010/04RM).

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

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Smythe HA, Moldofsky H. Two contributions to understanding of the ''fibrositis'' syndrome. Bull Rheum Dis 1977;28:928–31.

2 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.

3 Wolfe F, Smythe HA, Yunus MB *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.

4 Wolfe F, Clauw DJ, Fitzcharles M-A *et al*. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res 2010;62:600–10.

5 Salaffi F, Sarzi-Puttini P. Old and new criteria for the classification and diagnosis of fibromyalgia: comparison and evaluation. Clin Exp Rheumatol 2012;30(Suppl 74):3–9.

6 Fitzcharles MA, Boulos P. Inaccuracy in the diagnosis of fibromyalgia syndrome: analysis of referrals. Rheumatology 2003;42:263–7.

7 Wolfe F, Brähler E, Hinz A, Häuser W. Fibromyalgia prevalence, somatic symptom reporting, and the dimensionality of polysymptomatic distress: results from a survey of the general population. Arthritis Care Res 2013;65:777–85.

8 Silverman SL, Harnett J, Zlateva G, Mardekian J. Identifying fibromyalgia-associated symptoms and conditions from a clinical perspective: a step toward evaluating healthcare resource utilization in fibromyalgia. Pain Pract 2010;10:520–9.

9 Wolfe F, Clauw DJ, Fitzcharles M-A *et al*. 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.

10 Usui C, Hatta K, Aratani S *et al*. The Japanese version of the 2010 American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and the Fibromyalgia Symptom Scale: reliability and validity. Mod Rheumatol 2012;22:40–4.

11 Fitzcharles M-A, Ste-Marie PA, Panopalis P *et al*. 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.

12 Bidari A, Hassanzadeh M, Ghavidel Parsa B *et al*. Validation of the 2010 American College of Rheumatology preliminary diagnostic criteria for fibromyalgia in an Iranian population. Rheumatol Int 2013;33:2999–3007.

13 Usui C, Hatta K, Aratani S *et al*. 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.

14 Folstein MF, Folstein SE, McHugh PR. ''Mini-mental state'': a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

15 Bennett RM, Friend R, Jones KD *et al*. The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties. Arthritis Res Ther 2009;11:R120.

16 Luciano J V, Aguado J, Serrano-Blanco A, Calandre EP, Rodriguez-Lopez CM. Dimensionality, reliability, and validity of the revised fibromyalgia impact questionnaire in two spanish samples. Arthritis Care Res 2013;65:1682–9.

17 McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40–66.

18 Alonso J, Prieto L, Antó JM. [The Spanish version of the SF-36 Health Survey (the SF-36 health questionnaire): an instrument for measuring clinical results]. Med Clin (Barc) 1995;104:771–6.

19 Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315–25.

20 Munguía-Izquierdo D, Segura-Jiménez V, Camiletti-Moirón D *et al*. Multidimensional fatigue inventory: Spanish adaptation and psychometric properties for fibromyalgia patients. The Al-Andalus study. Clin Exp Rheumatol 2012;30(Suppl 74):94–102.

21 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–71.

22 Wiebe JS, Penley JA. A psychometric comparison of the Beck Depression Inventory-II in English and Spanish. Psychol Assess 2005;17:481–5.

23 Lobo A, Ezquerra J, Gómez Burgada F, Sala JM, Seva Díaz A. [Cognocitive mini-test (a simple practical test to detect intellectual changes in medical patients)]. Actas Luso Esp Neurol Psiquiatr Cienc Afines 1979;7:189–202.

24 Martin AA, Marzo PF. Delta: a new measure of agreement between two raters. Br J Math Stat Psychol 2004;57(Pt 1):1–19.

25 McBeth J, Jones K. Epidemiology of chronic musculoskeletal pain. Best Pract Res Clin Rheumatol 2007;21:403–25.

26 Queiroz LP. Worldwide epidemiology of fibromyalgia. Curr Pain Headache Rep 2013;17:356.

27 Mas AJ, Carmona L, Valverde M, Ribas B. Prevalence and impact of fibromyalgia on function and quality of life in individuals from the general population: results from a nationwide study in Spain. Clin Exp Rheumatol 2008;26:519–26.

28 Henriksen M, Lund H, Christensen R *et al*. 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.

29 Aparicio VA, Carbonell-Baeza A, Ortega FB *et al*. Usefulness of tenderness to characterise fibromyalgia severity in women. Clin Exp Rheumatol 2011;29(Suppl 69):S28–33.

30 Yunus MB, Aldag JC. The concept of incomplete fibromyalgia syndrome: comparison of incomplete fibromyalgia syndrome with fibromyalgia syndrome by 1990 ACR classification criteria and its implications for newer criteria and clinical practice. J Clin Rheumatol 2012;18:71–5.