JCR: Journal of Clinical Rheumatology

Psychometric properties of the Fibromyalgia Survey Questionnaire (FSQ) in Chilean women with fibromyalgia

--Manuscript Draft--

Manuscript Number:	JCR-20-280R2
Full Title:	Psychometric properties of the Fibromyalgia Survey Questionnaire (FSQ) in Chilean women with fibromyalgia
Article Type:	Original Articles
Keywords:	Fibromyalgia; illness index severity; musculoskeletal pain; outcome assessment.
Corresponding Author:	Lidia Gómez Pérez, phD Pontificia Universidad Catolica de Chile Santiago, Región Metropolitana CHILE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Pontificia Universidad Catolica de Chile
Corresponding Author's Secondary Institution:	
First Author:	Carla Aguirre Cárdenas, BA
First Author Secondary Information:	
Order of Authors:	Carla Aguirre Cárdenas, BA
	María Cecilia Oñederra, MA
	Catalina Esparza Benavente, BA
	Josefina Durán Santa Cruz, MD
	Matías Gonzalez Tuga, PhD
	Lidia Gómez Pérez, phD
Order of Authors Secondary Information:	
Manuscript Region of Origin:	CHILE
Abstract:	Objective: to evaluate the psychometric properties of the Chilean version of the FSQ. Methods: Women with fibromyalgia (FM, n = 214), women with rheumatoid arthritis (RA, n = 97), and women without chronic pain (attended at the Gynecologist, G, n = 117) from the Red Salud UC-Christus (Santiago, Chile) participated. Women with FM completed the Fibromyalgia Survey Questionnaire ; Fibromyalgia Impact Questionnaire Revised ; Numerical Pain Rating Scale; Pain Catastrophizing Scale; Pain Vigilance and Awareness Questionnaire; Patient Health Questionnaire-15; and Short-Form Health Survey . Two weeks later, they completed the FSQ again by phone (n=120). Results: the FSQ total scale showed excellent to good internal consistency at T1 (α = .91, ω = .91) and T2 (α = .78, ω = .78) and good test-retest reliability (ICC=.79; 95%CI: .72-.85). It showed medium to large correlations with the other measures. Discriminant analysis between FM group and Control group (RA and G) revealed that the FSQ total scale reached a classification accuracy of 81.3%. ROC curve (AAUC= .88; 95%CI: .85-.92) showed that the best FSQ cutoff was 17, resulting in sensitivity 89% (95%CI:.84-.93) and specificity 75% (95%CI:.69-.80). Considering the FM diagnosis performed by a rheumatologist as the gold standard, sensitivity and specificity of the modified 2010 ACR preliminary criteria for FM were 92.8% (95%CI:.88-.96) and 63.4% (95%CI:.57-.70), respectively. Conclusion: the Chilean version of the FSQ presents good psychometric properties and is a useful tool in clinical settings to assist in FM diagnosis and symptomatology assessment. A cutoff score of ≥ 17 appears to be the most appropriated for Chilean population.

Santiago 17 June 2020

Dear Editor and Reviewer,

Thank you very much again for your comments. These are really well appreciated. We have followed your suggestions and respond to the comments below:

Reviewer #1:

Comment 1. The authors have done an excellent job responding to the reviewers' concerns. I particularly appreciated the detail in regard to the conduct of the power analysis and wondered why at least a short version of their power analysis was not added to the paper. These data lend support to the rigor of the methodology.

Response 1: Thank you very much for the positive comments. As recommended, we have included the following paragraph with the power analyses in the data analyses section, after describing the analyses conducted (Page 12, line 249):

"*Sample size calculation. The sample size needed for the abovementioned analyses was calculated a priori. First, we used the program EPIDAT 4.2. (50) to determine the sample needed to calculate sensitivity and specificity. For that, we used the values reported by Wolfe et al (11), that is 96,6% of sensitivity and 91.8% of specificity, with a ratio of sick vs no sick of 1, an absolute precision of 5% and a confidence interval of 95%. The results of these analyses indicated that we needed 233 participants in total (117 in the fibromyalgia group and 116 in the non-fibromyalgia group). Then we used the software GPower 3.1. (51) to calculate the sample size needed for the discriminant analyses, using the following parameters: a medium size effect of f2V =.0625, with an alpha level = .05, and power = .80. The results indicated that we needed a total sample of 128. Then, we used the program MedCalc (52) to calculate how many participants we needed to conduct the Roc curve. The result of these analyses showed that for an alpha of .01 and a beta (probability of type two error) of .01, with an area under the curve of .70, a null hypothesis value of .05, and a ratio of sample size in negative/positive groups equal to 1, we required 182 participants in total, 91 in each group. The specificity and sensitivity analyses were the ones requiring a higher sample size (233 in total). Nonetheless, we recruited a larger sample of participants (the total sample size used for statistical analyses was 407, FM= 194, RA= 96, G=117). The reason was that we were simultaneously recruiting participants with research purposes other than the ones reported in the present manuscript. In this way we a priori assure we had the power we needed to perform our statistical analyses*".

In addition, we have added the following references to the references section:

Consellería de Sanidade, Xunta de Galicia, España; Organización Panamericana de la salud (OPS-OMS); Universidad CES, Colombia. *Epidat: programa para análisis epidemiológico de datos. Versión 4.2.* 2016. Retrieved from https://www.sergas.es/Saude-publica/EPIDAT?idioma=es

Faul F, Erdfelder E, Buchner A, & Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009; 41: 1149-1160.

Schoonjans F. *Medcalc statistics for biomedical research: software manual.* Belgium: MedCalc;1998.

Associate editor comment: Please incorporate a paragraph in the main text with power analysis.

> **Response:** As recommended, we have included the paragraph with the power analyses in page 12, line 249. We have also added the references of the programs used for calculating the sample size in the reference list.

Thank you very much again. Sincerely,

Lydia Gómez Pérez

Psychometric properties of the Fibromyalgia Survey Questionnaire (FSQ) in Chilean women with fibromyalgia

Short running head: Chilean Validation FSQ.

Carla Aguirre Cárdenas[a*], B.A.; Maria Cecilia Oñederra[a], M.A.; Catalina Esparza Benavente[a], M.A.; Josefina Durán Santa Cruz[b], MD; Matías González Tugas[c], PhD.; & Lydia Gómez-Pérez, PhD.

(*) The first and the second authors equally contributed to the present manuscript.

[a] Escuela de Psicología, Facultad de Ciencias Sociales, Pontificia Universidad Católica de Chile.

[b] Departamento de Inmunología Clínica y Reumatología, Escuela de Medicina, Pontificia Universidad Católica de Chile.

[c] Departamento de Psiquiatría, Escuela de Medicina, Pontificia Universidad Católica de Chile.

Corresponding autor: Lydia Gómez Perez. Email: lgomeze@uc.cl; lydiagp2@gmail.com. Escuela de Psicología. Facultad de Ciencias Sociales, Pontificia Universidad Católica de Chile, Campus San Joaquín. Avda. Vicuña Mackenna 4860, Macul. Santiago, Chile.

Funding: This study is supported by the Chilean National Commission for Scientific and Technological Research (CONYCIT, Comisión Nacional de Investigación Científica y Tecnológica), FONIS Regular SA16I0305 (Lydia Gómez-Pérez, Principal Investigator). This study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Psychometric properties of the Fibromyalgia Survey Questionnaire (FSQ) in Chilean women with fibromyalgia

Abstract

Objective: to evaluate the psychometric properties of the Chilean version of the FSQ. **Methods:** Women with fibromyalgia (FM, *n* = 214), women with rheumatoid arthritis (RA, *n*= 97), and women without chronic pain (attended at the Gynecologist, G, *n* = 117) from the *Red Salud UC-Christus* (Santiago, Chile) participated. Women with FM completed the *Fibromyalgia Survey Questionnaire*; *Fibromyalgia Impact Questionnaire Revised*; *Numerical Pain Rating Scale; Pain Catastrophizing Scale; Pain Vigilance and Awareness Questionnaire; Patient Health Questionnaire-15; and Short-Form Health Survey*. Two weeks later, they completed the FSQ again by phone (n=120). **Results:** the FSQ total scale showed excellent to good internal consistency at T1 (α= .91, ω= .91) and T2 (α= .78, ω= .78) and good test-retest reliability (ICC=.79; 95%CI: .72-.85). It showed medium to large correlations with the other measures. Discriminant analysis between FM group and Control group (RA and G) revealed that the FSQ total scale reached a classification accuracy of 81.3%. ROC curve (AAUC= .88; 95%CI: .85-.92) showed that the best FSQ cutoff was 17, resulting in sensitivity 89% (95%CI:.84-.93) and specificity 75% (95%CI:.69-.80). Considering the FM diagnosis performed by a rheumatologist as the gold standard, sensitivity and specificity of the modified 2010 ACR preliminary criteria for FM were 92.8% (95%CI:.88-.96) and 63.4% (95%CI:.57-.70), respectively. **Conclusion:** the Chilean version of the FSQ presents good psychometric properties and is a useful tool in clinical settings to assist in FM diagnosis and symptomatology assessment. A cutoff score of ≥ 17 appears to be the most appropriated for Chilean population.

Key Indexing Terms: fibromyalgia; illness index severity; musculoskeletal pain; outcome assessment.

Introduction

Fibromyalgia (FM) is a disorder, characterized by chronic, generalized musculoskeletal pain, and the presence of other psychosomatic, cognitive, and neurological symptoms (1-3). It may be difficult to diagnose, given that it is highly comorbid with several rheumatic, neurological, and psychiatric conditions (4-7) and that its symptoms overlap with the symptoms of these conditions (4, 8-9). In 2010, the American College of Rheumatology (ACR) adapted the FM diagnostic criteria (3) by changing the focus from physical pain, to FM's multidimensional nature. Based on these changes, the Fibromyalgia Survey Questionnaire (FSQ) was developed, which has shown to be useful to diagnose FM in epidemiological and clinical studies (10-11). The FSQ consists of two scales: The Widespread Pain Index (WPI) and the Symptom Severity Scale (SS). The WPI consists of a list of 19 non-articular regions in which patients indicate whether they have had pain or tenderness. WPI scores range from 0-19. The SS assesses the extent and intensity in which patients experience six core symptoms: sleep disturbance, fatigue, cognitive problems, headaches, abdominal pain or cramps, and depression. SS scores range from 0 to 12. The sum of WPI and SS constitutes the FSQ total Scale (also referred as the Polysymptomatic Distress Scale, PSD), which not only allows establishing a FM diagnosis (Score 0-31; cut-off score ≥ 13), but also assessing the intensity and symptomatic severity of the disorder (3, 11-13). Thus, FM can be studied on a dimensional or continuum scale, allowing further exploration of the syndrome (8, 14-16). FM diagnosis is made when the following conditions are met: 1) Generalized pain, defined as pain in at least 4 or 5 regions, is present; 2) Symptoms have been present at a similar level for at least 3 months; 3) WPI ≥ 7

and SS ≥ 5 or WPI= 4-6 and SS ≥ 9. Currently, a FM diagnosis is valid irrespective of other diagnoses, and it does not exclude the presence of other illnesses (17); however, we started participants' recruitment before this criterion was stablished and, therefore, participants with other illnesses were excluded in the present research. Conditions 1, 2, and 3 can be assessed by FSQ (11), which works as a screening tool. Nonetheless, in order to confirm the diagnosis, differential diagnosis performed by a specialist is needed. Therefore, the FSQ is not meant to replace or remove the use of other resources or diagnostic criteria, but rather, it can be understood as an alternative method to support early diagnosis and treatment of FM in various clinical contexts, and at a longitudinal level (3,11).

So far, the FSQ has been validated in Canadian (18), French (19), American (20), Japanese (21), Iranian (22-23), Turkish (24), and Spanish populations (25). It has shown acceptable levels of sensitivity (with ranges from 64% to 96%), specificity (ranging from 60% to 100%), reliability (with α ranges between .60 and .85), and validity (significant correlation coefficients with different questionnaires, such as FIQ-R, PHQ, and SF-12).

Despite the fact that the FSQ has shown to be useful to diagnose FM in epidemiological and clinical studies, and have been validated in many countries, no previous studies have examined the psychometric properties of the FSQ in Chile. Furthermore, cut-off scores for Chilean population are not available yet. Therefore, in our study, we aimed to evaluate the psychometric properties of the Chilean version of the FSQ. Namely, we assessed its reliability, (internal consistency and test-retest reliability), and discriminant validity. We also calculated the best FSQ cutoff score for this population and assessed its sensitivity and specificity.

Materials and Methods

Study design and participants

Adult women with FM (n = 214) and women with Rheumatoid Arthritis (RA, n= 97) that attended *Red Salud UC-Christus* (Santiago, Chile) participated in the present study. Additionally, adult women without chronic pain attending the Gynecology Unit of the same institution participated (G, n = 117). Exclusion criteria were being pregnant, presence of cancer, systemic vascular disease, mild to severe neurological problems, inflammatory rheumatic diseases, and comorbidity between FM and RA. A total of 431 subjects were recruited in the study. From this sample, 19 FM were excluded due to exclusion criteria, and 5 (FM = 1, RA= 1, G= 3) were excluded due to FSQ not being completed. As a result, a total sample of 407 (FM= 194, RA= 96, G=117) subjects were used for statistical analysis.

Instruments

Fibromyalgia Survey Questionnaire (FSQ) (11). The FSQ assesses the main FM symptoms according to the 2010 ACR preliminary diagnostic criteria (3). The elaboration of this instrument implied the modification of these preliminary criteria so that the diagnosis can be based on the symptoms self-reported by the patients (11). The Spanish version of the test (25) was adapted to be used in Chilean population.

Fibromyalgia Impact Questionnaire, revised version (FIQ-R) (26). The FIQ-R comprises 21 items in a 0-10 rating scale, which assesses physical functioning and FM symptom severity. It has three subscales (Functioning, Pain Impact, and Symptom Severity). Functioning scale includes nine items, assessing difficulty performing different daily activities. Pain impact scale consists of two items related to FM general impact. Finally, Symptom Severity scale has 10 items involving symptoms commonly reported by FM patients. A total score of the questionnaire can also be computed. The Spanish version of this questionnaire possesses adequate psychometric properties (27). We assessed whether this version of the

questionnaire was considered appropriate by Chilean patients using the sample of the present research (see method section). The Chilean version of this instrument showed adequate construct validity, internal consistence (ω= .96 for the total score) and test-retest reliability (ICC = .90) (28).

Numerical Pain Rating Scale (NPRS) (29). *It* consists of four numeric scales ranging from 0-10, assessing pain intensity at the time of the interview, as well as the worst, the slightest and the average pain experienced during the week.

Pain Catastrophizing Scale (PCS) (30). It is the most frequently used instrument to assess pain-related catastrophic thoughts and emotions. It includes 13 self-report items in a Likert scale (ranging from zero to four), comprising three subscales: Rumination, Magnification, and Hopelessness. A total score can also be calculated. Its Spanish version possesses adequate psychometric properties (31). Furthermore, it has been previously validated in Chile in two samples of pregnant women (32, 33). In the present sample, we assessed whether the previous Chilean version of the questionnaire was considered appropriate by Chilean patients with FM (see method section). Further analyses conducted with this sample indicated that this instrument showed appropriated construct validity, internal consistency (Cronbach alpha was .94 for total PCS score), and test-retest reliability (ICC was .84) (34).

The Pain Vigilance and Awareness Questionnaire (PVAQ) (35). This instrument evaluates pain hypervigilance. It consists of nine items rated from zero to five and organized in two subscales: Active Vigilance and Passive Awareness. The psychometric properties of this scale are adequate. A total score of the questionnaire can also be calculated. We also assessed whether the Spanish version of this questionnaire was considered appropriate by Chilean patients using the present sample (see method section). Preliminary analyses of the Chilean

version of this instrument showed acceptable construct validity, internal consistency (Cronbach alpha was .78 for the total score), and test retest reliability (ICC = .83) (36).

Patient Health Questionnaire-15 (PHQ-15) (37). The PHQ-15 is a somatic symptom subscale that assesses the presence of physical problems during the last 4 weeks. It comprises 15 items, rated from zero-two. Patients can be classified in four categories according to symptom severity: (scores = 0–4), low (scores = 5–9), medium (scores = 10–14), and high (scores = 15–30). The Spanish version was used (38), which has adequate psychometric properties. We also assessed whether the Spanish version of this questionnaire was considered appropriate by Chilean patients using the present sample (see method section). In the present sample, the internal consistency of the PHQ-15 was ω= .78.

Short-Form Health Survey (SF-12) (39). It is a quality of life questionnaire which appraises degree of well-being and functional ability. It is a shorter version of the SF-36. It consists of 12 items related to eight dimensions (Physical Function, Social Function, Physical Role, Emotional Role, Mental Health, Vitality, Bodily Pain, and General Health). We used the Chilean version of this questionnaire, which possess good psychometric properties (40). In the present sample, the internal consistency of the SF-12 was ω= .86.

Brief Pain Inventory Interference subscale (BPI-IS) (41). It assesses the degree of pain interference in daily life. It comprises seven items rated from zero to ten, assessing different areas: General Activity, Mood, Walking Ability, Normal Work, Relationships with other people, Sleep, and Enjoyment of Life. In the present study, we used the Spanish version of this questionnaire, which have shown adequate reliability and validity (42). Whether the Spanish version of this questionnaire was considered appropriate by Chilean patients was also

examined using the present sample (see method section). In the present sample, the internal consistency of this subscale was ω= .91.

Patient Health Questionnaire-9 (PHQ-9) (43). It assesses the presence of depressive symptoms in the last 2 weeks, according to the DSM-IV criteria. It consists of nine items, rating level of severity (zero to three). According to the obtained score and referred symptoms, diagnostic categories can be identified, as well as symptomatology type and severity, including: Major Depressive Syndrome, Other Depressive Syndrome, Positive Depressive Symptoms, and Negative Depressive Symptoms. The Chilean validated version was used in this study (44). In the present sample, the internal consistency of the PHQ-9 was ω= .84.

Procedure

This study was approved by the Medicine Scientific Ethics Committee from the Pontificia Universidad Católica de Chile (protocol number 16-210). In order to assess whether the Spanish version of the questionnaire was considered appropriate by Chilean patients, before collecting the study sample, this version of the scale, as well as other instruments not yet validated in Chile at the moment of the study in patients with FM (namely the PCS, FIQ-R, the PVAQ, the PHQ-15, and the BPI-IS) were applied to ten Chilean women with FM that were attending the rheumatology unit of the Centro Médico San Joaquín, which is part of Red Salud UC Christus. They were asked to indicate whether the items of these questionnaires were comprehensible, and the wording was clear to Chilean speakers. They could openly suggest any changes in the wording they deemed necessary to improve the scale. All items were considered comprehensible for all the participants. Similarly, they indicated that the items were worded in a way that they were adequate for its use in Chilean population, and no

changes were suggested. We did not consider these data in the statistical analyses conducted in order to assess the psychometric properties of the FSQ.

Participants with FM and participants with RA were invited to participate in the study by a research assistant when they were waiting to be attended by their rheumatologist. Additionally, FM cases were identified using a computer-based system that sent weekly notices of subjects with a diagnosis of fibromyalgia in electronic clinical notes. Furthermore, flyers and banners were placed in the rheumatology clinic waiting room with contact information when a research assistant was not present. We confirmed participants diagnoses by asking participants' physicians. Furthermore, we specifically asked the physicians to double-check that those patients with FM did not present RA, and vice versa. If the doctor was not available to confirm her patients diagnoses, we verified it using clinical records. Women without chronic pain were invited to participate while they were waiting to be attended by their gynecologist. The research assistant explained the purpose of the study and obtained written informed consent to participate as well as to publish the results of the study. Then, she asked participants to complete the questionnaires. Participants with FM completed the full battery of questionnaires, while participants with RA and participants without chronic pain only completed the FSQ, the NPRS, and some questions regarding clinical and sociodemographic characteristics. When a participant reported having vision or reading problems that prevented them to fill the questionnaires, they received help from the research assistant to complete them. In order to examine test-retest reliability of the FSQ, we asked participants with FM to complete this questionnaire again, this time by phone, two weeks later.

Statistical analysis

Psychometric properties of the FSQ were assessed using R-Statistical version 3.5.1 and SPSS version 24. Before conducting the main analyses, we examined whether there were differences between the groups regarding the sociodemographic variables (age, region, marital status, educational level, and employment status). We conducted a non-parametric analysis (Kruskal Wallis Test) to check for differences in age, because the groups differed in sample size; and Chi-squared Tests to examine differences between the groups in the categorical variables. Then, we performed the following analyses:

Reliability

Internal consistency. We computed Cronbach's alpha (α) and McDonald's omega coefficient (ω) for the SS and the FSQ total scales at T1 and T2, using all the participants of the study. We considered scores above .70 for alpha and omega as satisfactory (45).

Test–Retest reliability. We calculated the intra-class correlation coefficients (ICCs) between T1 and T2 of the SS, the WPI, and the FSQ total scales, using only the sample of participants with fibromyalgia. We used a 95% confidence interval, based on a single-rating, absolute-agreement, two-way mixed-effects model. We used the following ICC values range: poor (<.50), moderate (.50-.75), good (.75-.90), and excellent agreement (>.90) (46). For this data analysis, only subjects who participated in the follow-up [n=120, 61.9%] were considered.

Validity

In order to assess convergent validity, we calculated Pearson's correlations between the FSQ total scale in FM group and the additional instruments (FIQ-R, NPRS, PCS, PVAQ, PHQ-15, SF-15, BPI, and PHQ-9). In order to assess discriminant validity, we conducted two hierarchical discriminant analyses. The first one considering the score in the FSQ total scale as the

independent variable and predicting whether participants belonged to the FM group or the Control group (RA + G). The second one considering only the FM and the AR groups and examining whether the FSQ could discriminate between these samples. In both analyses, we introduced age, educational level, and marital and employment status as control variables in the first step of the model. Then, we introduced the FSQ total score in the second step and examined whether the accuracy of the model was improved by the inclusion of this variable, once the covariables had been controlled for. Before conducting the analyses, categorical variables with more than two categories were transformed into dummy variables. For marital status the baseline category was "Married/Live-in partner"; for educational level the baseline category was "high-school education or less", and for employment status the baseline category was "being working" (see Table 1). Following Huberty's suggestions (47) we calculated the discriminant function with half of the sample, and then cross-validated these results with the other half of the sample. The levels of accuracy reported are the ones calculated during the cross-validation phase.

Cut-off score, sensitivity, and specificity

We conducted receiver operating characteristic curve (ROC, 48) in order to determine the Area Under the Curve (AUC) and the best FSQ total scale cut-off score. In order to calculate the FSQ total scale Adjusted AUC (AAUC), we conducted Covariate-Adjusted ROC (49) controlling for age, educational level, and marital and employment status. Finally, we examined the sensitivity and specificity for the FSQ modified 2010 preliminary criteria for FM recommended by Wolfe et al. (11), that is presenting a WPI score ≥ 7 and a SS score ≥ 5 or presenting a WPI score equal to 3-6 and a SS score ≥ 9 , as well as presenting these symptoms for at least three months.

Sample size calculation. The sample size needed for the abovementioned analyses was calculated a priori. First, we used the program EPIDAT 4.2. (50) to determine the sample needed to calculate sensitivity and specificity. For that, we used the values reported by Wolfe et al (11), that is 96,6% of sensitivity and 91.8% of specificity, with a ratio of sick vs no sick of 1, an absolute precision of 5% and a confidence interval of 95%. The results of these analyses indicated that we needed 233 participants in total (117 in the fibromyalgia group and 116 in the non-fibromyalgia group). Then we used the software GPower 3.1. (51) to calculate the sample size needed for the discriminant analyses, using the following parameters: a medium size effect of f2V =.0625, with an alpha level = .05, and power = .80. The results indicated that we needed a total sample of 128. Then, we used the program MedCalc (52) to calculate how many participants we needed to conduct the Roc curve. The result of these analyses showed that for an alpha of .01 and a beta (probability of type two error) of .01, with an area under the curve of .70, a null hypothesis value of .05, and a ratio of sample size in negative/positive groups equal to 1, we required 182 participants in total, 91 in each group. The specificity and sensitivity analyses were the ones requiring a higher sample size (233 in total). Nonetheless, we recruited a larger sample of participants (the total sample size used for statistical analyses was 407, FM= 194, RA= 96, G=117). The reason was that we were simultaneously recruiting participants with research purposes other than the ones reported in the present manuscript. In this way we a priori assure we had the power we needed to perform our statistical analyses.

Results

We present the demographics and clinical characteristics of the sample in Table 1 and 2. The descriptive statistics of each instruments for each group are described in Table 3. The sample was mainly composed of middle age women, from Chile's capital Santiago (84.3%),

predominantly married/live-in partner (58%), with a full/part time employment (50.9%) and a high school education level or less (36.6%).

Significant differences were found between the groups in the mean age of the participants, RA being the oldest group and G the youngest one. The groups also significantly differed in marital status, educational level, and employment status (See Table 1).

Participants with FM presented an average of 7.46 years with pain and 3.61 years since their diagnosis. They had a significantly higher score in the FSQ total scale (F[2, 404] =181.99, p <.0001), Symptom Severity Scale (SS) (F[2, 404] =86.01, p <.0001), Widespread Pain Scale (F[2, 404]=177.86, p <.0001) and consumption of antidepressants (92.5%) than other groups (7.5%) (See Table 2). The most frequently used antidepressant type was serotonin and norepinephrine reuptake inhibitors (SNRIs, 67.9%).

Reliability.

Internal consistency reliability. FSQ total scale indicates excellent to good internal consistency at T1 (α= .91, ω= .91) and T2 (α= .78, ω= .78). However, SS only shows good internal consistency at T1 (α= .74, ω= .76) but a questionable internal consistency at T2 (α= .60, ω= .62).

Test-Retest reliability. The SS showed moderate test-retest reliability (ICC = .65 [.53- .74]), while the WPI (ICC= .75 [.66-.82]) and PSD (ICC=.79 [.72-.85]) showed good test-retest reliability.

Validity.

Convergent validity. Correlations between PSD and other measures were mostly medium to large and statistically significant (See Table 3, SDC 3, descriptive statistics of the instruments used). There were large and positive correlations with the PHQ-15 (r=.62, p<.0001), the FIQ-

R (r=.60, p<.0001), and the NPRS (r=.51, p<.0001). The PSD showed medium and positive correlations with the PHQ-9 (r=.49, p<.0001), the BPI-PI (r=.47, p<.0001), and the PCS (r=.31, p<.0001); and an inverse and medium association with the SF-12 (r= -.46, p<.0001). Finally, related to PVAQ, it presented a small and positive correlation (r=.22, p=.002).

Discriminant validity. The results of the hierarchical discriminant analysis conducted to predict whether participants belong to the FM or Control group (RA+G) showed that there were significant mean differences between the groups in the predictors 'FSQ total score' (Wilk's λ =.612, p<.0001) and 'being widowed' (Wilk's λ =.963, p=.006). In the first step, when only the control variables were included in the analyses, the linear equation significantly predicted group membership (Wilk's λ = .883, 2 (11) = 24.36, p=.011), accounting for 11.6% of between-group variability. The relative contribution of each of the control variables is reported in the structure matrix presented in Table 4. A loading higher than .30 is considered as the cut-off between important and less important variables (53). The variable with the higher contribution to the equation was being married/live-in partner versus being widowed, followed by being working versus being a student, being working versus being retired, being working versus being in medical leave, having high-school education or less versus having technical studies, and being married/live-in partner versus being single. Namely, being married/live-in partner (when compared with those widowed and single), being a student (instead of a worker), and being working (instead of retired) were variables inversely associated with having FM; whereas being working (versus in medical leave) and having a high-school education or less were directly associated with having FM. The classification accuracy was 56.7%. When the FSQ total score was introduced in the second step, the model significantly improved (Wilk's λ = .561, 2 (12) = 113.37, p <.0001)

accounting for a higher between-group variability (44%) and showing a classification accuracy of 81.3%. The relative contribution of each of the control variables is reported in the structure matrix presented in Table 4. Only the FSQ total score significantly contributed to the linear equation (*r* =.90).

The results of the hierarchical discriminant analysis conducted to predict whether participants belonged to the FM or the AR group showed that there were significant differences between the group in the following predictors: mean FSQ total score (Wilk's λ=.721, *p*<.0001), being married/living-in partner versus being widowed (Wilk's λ=.923, *p*=.001), being working versus retired (Wilk's λ=.952, p=.008), age (Wilk's λ=.955, p=.010), having high-school education or less versus technical school education (Wilk's λ=.961, *p*=.018), and being married/living-in partner versus being single (Wilk's λ=.973, *p*=.048). In the first step, when only the control variables were included in the analyses, the linear equation significantly predicted group membership (Wilk's λ = .795, 2 (11) = 31.59, *p* =.001), accounting for 20.5% of the between group variability. The relative contribution of each of the control variables is reported in the structure matrix presented in Table 4. The predictors that presented a cutoff score higher than .30 were: being married/live-in partner versus being widowed, followed by being working versus retired, age, having high-school education or less versus technical school education, and being married/living-in partner versus being single, respectively. Specifically, the age, being married (when compared with those widowed or retired), being working (instead of retired) was inversely associated with being a member of the FM group; whereas having high-school education (versus technical school education) was directly associated with being a member of the FM group. The classification accuracy was 62.1%. When, the FSQ total score was introduced in the second

step, the model significantly improved (Wilk's λ = .573, 2 (12) =76.37, p<.0001) accounting for a higher between group variability (42.8%) and showing a significantly higher classification accuracy of 71.7%. The relative contribution of each of the control variables is reported in the structure matrix presented in Table 4. Only the FSQ total score significantly contributed to the linear equation (r =.72).

Cut-off score, sensitivity, and reliability. ROC curves showed that the FSQ total scale as a predictor for the classification to the groups (FM and control [RA + G]) had a very good performance to discriminate the presence versus absence of fibromyalgia (AUC= .88; 95%CI: .84-.91), and a Covariate Adjusted ROC showed an almost identical result, (AAUC= .88 :95%CI: .85-.92) (see Figure 1). The best cutoff that maximizes (sensitivity + specificity) is 17 (see Table 5), resulting in sensitivity .89 (95% CI: .84-.93), specificity .75 (95% CI: .69-.80), Positive Predictive Value (PPV) .76, Negative Predictive Value (NPV) .88, Positive Likelihood Ratio (PLR), 3.50, and Negative Likelihood Ratio (NLR) .15.

When using the modified 2010 ACR preliminary criteria for fibromyalgia (11) to classify the patients, instead of the FSQ total score, the level of sensitivity was high (92.8%; 95% CI:.88-.96) and the level of specificity was adequate (63.4%; 95% CI: .57-.70). Nonetheless, FSQ specificity was higher in G group (77.8%; 95% CI: .69-.85) than the RA group (45.8%; 95% CI: .36-.56).

Discussion

The present study showed acceptable reliability and validity for the FSQ as a fibromyalgia diagnostic instrument. PSD (total scale) and WPI presented good performance, especially regarding test-retest reliability, which indicated consistent answers through time. In addition, in line with our expectations, participants with FM presented higher SS, WPI and FSQ total

scale scores than participants in the control groups. Furthermore, the FSQ showed good convergent validity, as it presented positive and large correlations with the other measures. Mainly, the PSD was associated with somatic symptom severity, fibromyalgia impact, and pain intensity. Similar results were observed in previous validations (e.g. 19, 25), which provides more evidence of the impairment that physical and psychosocial symptoms of FM elicit on patients.

Despite this, it is important to consider the difference of results between the scales, as SS did not perform as well as the others scales, decreasing in .14 its internal consistency at T2 [α= .60, ω= .62] and getting a moderate test-retest reliability [ICC .62]. One explanation may be that some of the participants were first interviewed after receiving their first consultation for FM with a rheumatologist, and then interviewed again two weeks after starting a new medical treatment. Therefore, symptoms' improvement may be responsible for the lower internal consistency at T2 and the moderate test-retest reliability. Furthermore, the first rheumatologist assessment itself (and not only the medical treatment) can produce an initial effect in the symptoms, which could be affecting the stability of the SS scores. An alternative explanation is that nine of the 12 scores of the SS explore symptoms experienced during the last week, which usually fluctuate overtime. Although most measures show stability over time, SS variations at test-retest must be understood considering FM's dynamic nature.

When compared with our gold standard (rheumatologist diagnosis), the modified 2010 ACR preliminary criteria for FM (11) (WPI ≥ 7 and SS ≥ 5 or WPI 3–6 and SS ≥ 9) showed greater sensitivity (92.8) but lower specificity (63.4) than the FSQ total score (.89 and .75 respectively). The high sensitivity found in the modified 2010 ACR preliminary criteria and the FSQ total score shows that this instrument can correctly identify FM subjects. The FSQ

total score presented better specificity properties, and a better balance between sensitivity and specificity, than the modified 2010 ACR preliminary criteria (11).

The best cut-off score that maximizes sensitivity and specificity properties of the FSQ was 17 for the Chilean version, which is higher than the recommended by Wolfe et al. (11). Considering that patients with FM are a heterogeneous group, a possible explanation for the different cutoff scores can be cross-cultural differences, a trend that has also been reported by other studies (e.g. 21-22). It must also be considered that some longitudinal studies of patients with FM have shown fluctuations around the cutoff points (54-55). The presence of FM subgroups or clusters have been suggested, based on variations in pain and other symptomatology (14, 56), that might cause significant differences in the way patients experience and report the syndrome, especially in terms of psychosocial factors. As FM is still a highly complex condition, and diagnosis and assessment criteria keep being developed, FSQ has some limitations as well. For example, the fact that it is almost completely reliant on symptoms, and doesn't include the assessment of mechanistic factors (57-58), which could be a necessary change, as central sensitization is a core pathophysiological factor in FM diagnosis and assessment (2, 59-60).

FSQ is a fast self-administered, easy to understand questionnaire. Some advantages of our study are that this is the first validation of FSQ in Latin America. This is relevant given cultural differences may affect pain experiences. In addition, we studied an adequate sample size, controlled for differences in sociodemographic variables, and were able to follow up subjects, which is a critical aspect when assessing FM, as it is a dynamic condition with significant changes in symptomatology and presentation over time.

This study, however, presents some limitations. First, diagnosis was provided by the physicians attending the participants, and the research team did not conduct any standard healthcare professional examination as part of the research procedure. Nonetheless, we specifically asked the physicians to confirm that the patients with FM did not present RA, and vice versa. In spite of this, fibromyalgia could eventually be present among patients with RA, especially if we take into account that until recently, the diagnosis of FM was ruled out when the participant presented another inflammatory disease that could be otherwise explaining the pain (11). This could have impacted our cut-off score for fibromyalgia diagnosis and be partially explaining why our cut off score (≥17) was higher than the one previously reported in other countries (≥13). This potential co-ocurrence of FM and RA may have also affected the results of our discriminant analyses. Finally, this could also at least partially explain the low specificity found for the FSQ modified ACR 2010 preliminary diagnosis criteria recommended by Wolfe et al. (11) when using the participants with RA as the comparison group. Replicating our results in future studies in which this comorbidity is more rigorously controlled as well as in studies comparing the specificity and sensitivity of the FSQ using other chronic pain conditions as the comparative group is needed. Despite this limitation, this is the first study in which the validity of the FSQ is examined separating RA in a single group. In previous studies, the comparison non-fibromyalgia group generally comprises RA patients as well as patients with other pain disorders such as osteoarthritis (OA). However, the proportion of patients with noninflammatory chronic pain diagnoses such as OA (which frequently affects only a few joints and does not produce fatigue) is usually much higher than the proportion of patients with RA. Therefore, it is possible that in future studies using a comparison group comprising only patients with RA, the cutoff scores could be higher than the reported until

now in the literature. In this sense, our findings contribute to the literature by providing for the first time data regarding the validity of the FSQ to discriminate between patients with FM and patients with RA, and suggest that the ability of the test to discriminate between FM and RA may be lower than the ability to discriminate FM amongst other chronic pain diagnoses. Second, we considered the FM diagnosis performed by a physician as the gold standard and, therefore, our findings are subject to the same limitations and controversies that nowadays surround the FM diagnosis. In this sense, we agree with Galvez-Sánchez & Reyes del Paso (61) in that the current lack of clear consensus regarding the concept and diagnosis of FMS amongst medical professionals remains, and that complaints from health professionals and patients about the way the disease is diagnosed continue. Third, usually in studies analyzing test-retest reliability, not all the participants, but a subsample is contacted. As such, we did not invite all the study participants to complete the second assessment aimed to assess the FSQ test-retest reliability. Nonetheless, one limitation of the present research is that we did not adequately register how many participants were finally contacted for the test-retest and, therefore, we do not know which was the dropout rate for the second assessment. In case that the dropout was higher than 20%, the external validity of our results would have been compromised. However, although we do not have reasons to think that the dropout rate of our study was higher than the one presented in previous research, our results regarding the test-retest reliability of the FSQ must be taken with caution. In spite of this constraint, only two previous studies have examined the test-retest reliability of the FSQ (19, 24), and unlike our study, both of them have the limitation of having used Pearson´s correlations – instead of intraclass correlation – to assess this type of reliability, which is not appropriate (62). Therefore, despite our limitations, our study provides results about the reliability of the FSQ

that are not available in the current literature and overcomes the limitations of these two previous studies. Conducting more rigorous studies assessing the retest-retest reliability of the FSQ is in the future is indeed essential. As FM symptom variations are expected, it is important to have strong longitudinal reports to be able to capture the dynamics of the pain experience. Another limitation of the present study is that two of the questionnaires used to assess the FSQ convergent validity (i. e. the PHQ-15 and the interference subscale of the BPI) had been previously validated in Spain, but not in Chile. Furthermore, although the construct validity and reliability of two of the scales (the FIQ-R and the PVAQ) have been previously examined using the same sample as in the present research, the convergent validity of these two instruments remains to be conducted in Chile. Although these four instruments showed adequate reliability in the present sample, the results regarding the convergent validity using these questionnaires should be taken with caution and further studies about the convergent validity of these questionnaires need to be conducted. In addition, as the majority of the studies conducted about the validity of FSQ, participants were not community subjects, which might have limited the generalizability of our results. Moreover, we did not include patients from other settings, such as primary care. Furthermore, we recruited only female patients, so the results cannot be generalized to men with FM. Besides, discriminant analyses were performed using patients with RA and participants without chronic pain and, did not include patients with other diagnoses (such as osteoarthritis, spondyloarthropathies, polyneuropathy, etc.); therefore, the discriminant value of the test is limited only for patients with RA. Studies including participants with other pain diagnoses should be conducted in the future. It is also important to consider variations in T1 and T2 assessments, as they had

different conditions (e.g. T1 involved a face-to-face interview, while T2 was conducted over the phone).

In conclusion, the Chilean version of the FSQ has shown good psychometric properties. It is a useful tool in clinical settings to assist in FM diagnosis and symptomatology assessment. Setting a cutoff score of ≥ 17 appears to be the most effective approach in Chilean population.

Acknowledgment

We want to thank to all the women that participated in the present research as well as the reviewer of the paper for their time and insights. We also want to thanks to the research assistants Camila Belén Román Butrón, Débora Martellans, Alejandra Escobar Zuluaga, Ana Carolina de Obaldía, Alejandra Rivera, Claudia Toloza, Francisca Navarro, Javiera Soza, Laura Maldonado, Xaviera Acuña, Manuela Pacheco, and Montserrat Fellenberg, who contributed to the data collection. We also want to thank Dr. Jorge Flores, who helped edit the final version of the manuscript.

References

1. Arnold L, Clauw D, McCarberg B. Improving the Recognition and Diagnosis of Fibromyalgia. *Mayo Clin Proc.* 2011; 86:457-464.
2. Sarzi-Puttini P, Atzeni F, Masala I, et al. Are the ACR 2010 Diagnostic Criteria for Fibromyalgia Better than the 1990 Criteria? *Autoimmun Rev*. 2018; 17:33-35.
3. Wolfe F, Clauw D, Fitzcharles M, et al. The American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and Measurement of Symptom Severity. *Arthritis Care & Res*. 2010; 62:600-610.
4. Arnold L. The Pathophysiology, Diagnosis and Treatment of Fibromyalgia. *Psychiatr Clin North Am.* 2010; 33:375-408.
5. Bennett R. Diagnostic Criteria and Differential Diagnosis of the Fibromyalgia Syndrome. *J Musculoskelet Pain.* 2004; 12:59-64.
6. Natelson B. Fibromyalgia Syndrome: Diagnosis and Comorbidities. *CNS Spectr*. 2009; 16:8-11.
7. Russell I, Raphael K. Fibromyalgia Syndrome: Presentation, Diagnosis, Differential Diagnosis, and Vulnerability. *CNS Spectr.* 2008; 15:6-11.
8. Häuser W, Sarzi-Puttini P, Fitzcharles M. Fibromyalgia Syndrome: Under-, Over, and Misdiagnosis. *Clin Exp Rheumatol*. 2019; 116:90-97.
9. Valls-Llobet C. Diagnóstico Diferencial del Dolor y de la Fibromialgia. *Anu Psicología.* 2008; 39:87-92.
10. Häuser W, Wolfe F. Diagnosis and Diagnostic tests for Fibromyalgia (syndrome). *Reumatismo*. 2012; 64:194-205.
11. Wolfe F, Clauw D, Fitzcharles M, 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; 38:1113-1122.

12. Friend R, Bennett R. A Critical Examination of the Polysymptomatic Distress Scale Construct as a Symptom Severity Questionnaire. *J Rheumatol.* 2015; 42:1364-1367.
13. Wolfe F, Walitt B, Rasker J, et al. The Use of Polysymptomatic Distress Categories in the Evaluation of Fibromyalgia (FM) and FM Severity. *J Rheumatol*. 2015; 42:1-8.
14. Bidari A, Ghavidel-Parsa B, Ghalehbaghi B. Challenges in Fibromyalgia Diagnosis: From Meaning of Symptoms to Fibromyalgia Labeling. *Korean J Pain.* 2018; 31:147-154.
15. Wang S, Han C, Lee S, at al. Fibromyalgia Diagnosis: A Review of the Past, Present and Future. *Expert Rev Neurother.* 2015; 15:667-669.
16. Wolfe F, Brähler E, Hinz A, et al. Fibromyalgia Prevalence, Somatic Symptom Reporting, and the Dimensionality o Polysymptomatic Distress: Results from a Survey of the General Population. *Arthritis Care Res*. 2013; 65:777-785.
17. Wolfe F, Clauw D, Fitzcharles M, et al. Revisions to the 2010/2011 Fibromyalgia Diagnostic Criteria. *Semin Arthritis Rheum*. 2016; 46:319-329.
18. Ferrari R, Russell A. A Questionnaire Using the Modified 2010 American College of Rheumatology Criteria for Fibromyalgia: Specificity and Sensitivity in Clinical Practice. *J Rheumatol*. 2013; 40:1590-1595.
19. Fitzcharles M, Ste-Marie P, 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; 13:1-6.

20. Bennet R, Friend R, Marcus D, et al. Criteria for the Diagnosis of Fibromyalgia: Validation of the Modified 2010 Preliminary American College of Rheumatology Criteria and the Development of Alternative Criteria. *Arthritis Care Res.* 2014; 66:1364-1373.
21. 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* 2012; 23:846-850.
22. 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.
23. Bidari A, Ghavidel-Parsa B, Amir A, et al. Validation of Fibromyalgia Survey Questionnaire and Polysymptomatic Distress Scale in a Persian Population. *Rheumatol Int.* 2015; 35:2013-2019.
24. Yanmaz M, Atar S, Biçer M. The Reliability and Validity of the Turkish Version of Fibromyalgia Survey Diagnostic Criteria and Symptom Severity Scale. *J Back Musculoskel Rehabil.* 2016; 29:287-293.
25. Carrillo de la Peña M, Triñanes Y, González-Villar A, et al. Convergence Between the 1990 and 2010 ACR Diagnostic Criteria and Validation of the Spanish version of the Fibromyalgia Survey Questionnaire (FSQ). *Rheumatol Int.* 2014; 35:141-151.
26. Bennett R, Friend R, Jones K, et al. The Revised Fibromyalgia Impact Questionnaire (FIQR): Validation and Psychometric Properties. *Arthritis Res Ther.* 2009; 11:415.
27. Salgueiro M, García-Leiva J, Ballesteros J, et al. Validation of a Spanish Version of the Revised Fibromyalgia Impact Questionnaire (FIQR). *Health Qual Life Outcomes.* 2013;

1:132.

28. Escobar A, Román Butrón C, Vergés A, et al. *Psychometric properties of the Chilean version of the Fibromyalgia Impact Questionnaire Revised (FIQ-R) in a sample with Fibromyalgia: Preliminary results* [Conference presentation]. 15th International Congress of Behavioral Medicine (ICBM). Santiago, Chile. 2018, November 14-17.
29. Jensen, M, Turner, J, Romano, et al. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999; 83:157-162.
30. Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and Validation. *Psychol Asses.* 1995; 7:524-532.
31. García-Campayo J, Rodero B, Alda M, Sobradiel N, Montero J, Moreno S. Validación de la Versión Española de la Escala de Catastrofismo ante el Dolor (Pain Catastrophizing Scale) en la Fibromialgia. *Med Clin* 2008; 131:487-492.
32. Martínez M, Miró E, Sánchez A, et al. Spanish Version of the Pain Vigilance and Awareness Questionnaire: Psychometric Properties in a Sample of Women with Fibromyalgia. *Span J Psychol.* 2015; 17:1-8.
33. Esparza C, Soto M, Cumsille P, et al. *Validación de la Escala de catastrofismo ante el dolor en mujeres gestantes* [Conference presentation]. XIV Congreso Chileno de Psicología. Arica, Chile. 2019, November 13-15.
34. Vázquez-Taboada AR, Gómez-Pérez L, Torres-Flores MP et al. *Propiedades psicométricas de la Escala de Catastrofismo ante el Dolor (PCS) en una muestra de mujeres chilenas embarazadas* [Conference presentation]. Symposium Prediciendo la recuperación postparto at the VII Congreso Latinoamericano de Psicología de la Salud

591 [ALAPSA]. Cali, Columbia.2016, May 23-27.
592 35. Gómez-Pérez L, de Obaldía AC, Escobar A, et al. *Psychometric properties of the Chilean*
593 *version of the Pain Catastrophizing Scale in a sample of Chilean women with*
594 *fibromyalgia* [Conference presentation]. 15th International Congress of Behavioral
595 Medicine (ICBM). Santiago, Chile.2018, November 14-17.
596 36. Román C, Escobar A, Arrianza T, et al. *Psychometric properties of the Chilean version*
597 *of the Pain Vigilance and Awareness Questionnaire (PAVQ) in women with fibromyalgia*
598 [Conference presentation]. 15th International Congress of Behavioral Medicine
599 (ICBM). Santiago, Chile. 2018, November 14-17.
600 37. Kroenke K, Spitzer R, Williams J. The PHQ-15: Validity of a new Measure for
601 Evaluating the Severity of Somatic Symptoms. *Psychosom Med*. 2002; 64:258-266.
602 38. Ros S, Comas A, García-García M. Validación de la Versión Española del Cuestionario
603 PHQ-15 para la Evaluación de Síntomas Físicos en Pacientes con Trastornos de
604 Depresión y/o Ansiedad: Estudio DEPRE-SOMA. *Actas Esp Psiquiatr.* 2010; 38:345-
605 357.
606 39. Ware J, Kosinski, M, Keller S. A 12-item short-form Health Survey. Construction of
607 Scales and Preliminary Tests of Reliability and Validity. *Med Care*. 1996; 34:220-233.
608 40. Vera-Villarroel P, Silva J, Celis-Atenas K, et al. Evaluación del Cuestionario SF-12:
609 Verificación de la Utilidad de la Escala de Salud Mental. *Rev Med Chile.* 2014;
610 142:1275-1283.
611 41. Cleeland C, Ryan K. Pain Assessment: Global Use of the Brief Pain Inventory. *Ann Acad*
612 *Med Singapore*. 1994; 23:129-138.
613 42. De Andrés J, Cruces L, Canos M, et al. Validation of the Short Form of the Brief Pain

Inventory (BPI-SF) in Spanish Patients with Non-Cancer-Related Pain. *Pain Pract.* 2015; 15:643-653.

43. Kroenke K, Spitzer R, Williams J. The PHQ-9 Validity of a Brief Depression Severity Measure. *J Gen Intern Med.* 2001; 16:606-613.
44. Baader M, Molina J, Venezian S, et al. Validación y Utilidad de la Encuesta PH1-9 (Patient Health Questionnaire) en el Diagnóstico de Depresión en Pacientes Usuarios de Atención Primaria en Chile. *Rev Chil Neuro-Psiquiatr*. 2012; 50:10-22.
45. Nunnally J, Bernstein I. *Psychometric theory* (3rd ed.). New York: McGraw-Hill; 1994.
46. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016; 15:155-163.
47. Huberty CJ. *Applied Discriminant Analysis (1st ed.).* Michigan: Wiley-Interscience; 1994.
48. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* 2011; 12: 77.
49. Rodríguez-Álvarez MX, Vanda I. "*ROCnReg: An R Package for Receiver Operating Characteristic Curve Inference with and without Covariate Information*." arXiv2020: 2003.13111v2.
50. Consellería de Sanidade, Xunta de Galicia, España; Organización Panamericana de la salud (OPS-OMS); Universidad CES, Colombia. *Epidat: programa para análisis epidemiológico de datos. Versión 4.2.* 2016. Retrieved from https://www.sergas.es/Saude-publica/EPIDAT?idioma=es
51. Faul F, Erdfelder E, Buchner A, & Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009; 41: 1149-1160.

637 52. Schoonjans F. *Medcalc statistics for biomedical research: software manual.* Belgium:
638 MedCalc;1998.
639 53. Garson GD. *Factor analysis* (Statistical Associates Blue Book Series). Asheboro:
640 Statistical Associates Publishers; 2013.
641 54. Adams E, McElroy H, Udall M, et al. Progression of Fibromyalgia: Results from a 2-year
642 observational Fibromyalgia and Chronic Pain Study in the US. *J Pain Res*. 2016; 9:325-
643 336.
644 55. Wallit B, Fitzcharles M, Hassett A, et al. The Longitudinal Outcome of Fibromyalgia: A
645 Study of 1555 Patients. *J Rheumatol.* 2011; 38:2238-2246.
646 56. Bartley E, Robinson M, Staud R. Pain and Fatigue Variability Patterns Distinguish
647 Subgroups of Fibromyalgia Patients. *J Pain*. 2018; 19:372-381.
648 57. Egloff R, von Känel V, Müller U, et al. Implications of Proposed Fibromyalgia Criteria
649 Across Other Functional Pain Syndromes. *Scand J Rheumatol*. 2015; 44:416-424.
650 58. Wolfe F, Egloff N, Häuser W. Widespread Pain and Low Widespread Pain Index Scores
651 Among Fibromyalgia-positive Cases Assessed with the 2010/2011 Fibromyalgia
652 Criteria. *J Rheumatol.* 2016; 43:1743-1748.
653 59. Wolfe F, Butler S, Fitzcharles M, et al. Revised Chronic Widespread Pain Criteria:
654 Development from and Integration with Fibromyalgia Criteria. *Scand J Pain*. 2019;
655 18:77-86.
656 60. Ghavidel-Parsa B, Bidari A, Hajiabbasi A, et al. Fibromyalgia Diagnostic Model Derived
657 from Combination of American College of Rheumatology 1990 and 2011 Criteria.
658 *Korean J Pain.* 2019; 32:120-128.
659 61. Galvez-Sánchez CM, Reyes del Paso, GA. (2020). Diagnostic Criteria for Fibromyalgia:

Critical Review and Future Perspectives. *J Clin Med.* 2020;9: 1219.

62. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between two Methods of Clinical Measurement. *Lancet.* 1986; 1: 307-310.

Figure 1. Receiver operating characteristic curve (ROC) for the FSQ total scale as a predictor for the classification to the groups (FM and control [RA + G]). Area Under the Curve (AUC) = .88; 95%CI: .84-.91). Covariate Adjusted ROC (AAUC= .88 :95%CI: .85-.92). FM = Fibromyalgia; RA = Arthritis Rheumatoid; and G = Attended at the Gynecologist.

Psychometric properties of the Fibromyalgia Survey Questionnaire (FSQ) in Chilean women with fibromyalgia

Short running head: Chilean Validation FSQ.

Carla Aguirre Cárdenas[a*], B.A.; Maria Cecilia Oñederra[a], M.A.; Catalina Esparza Benavente[a], M.A.; Josefina Durán Santa Cruz[b], MD; Matías González Tugas[c], PhD.; & Lydia Gómez-Pérez, PhD.

(*) The first and the second authors equally contributed to the present manuscript.

[a] Escuela de Psicología, Facultad de Ciencias Sociales, Pontificia Universidad Católica de Chile.

[b] Departamento de Inmunología Clínica y Reumatología, Escuela de Medicina, Pontificia Universidad Católica de Chile.

[c] Departamento de Psiquiatría, Escuela de Medicina, Pontificia Universidad Católica de Chile.

Corresponding autor: Lydia Gómez Perez. Email: lgomeze@uc.cl; lydiagp2@gmail.com. Escuela de Psicología. Facultad de Ciencias Sociales, Pontificia Universidad Católica de Chile, Campus San Joaquín. Avda. Vicuña Mackenna 4860, Macul. Santiago, Chile.

Funding: This study is supported by the Chilean National Commission for Scientific and Technological Research (CONYCIT, Comisión Nacional de Investigación Científica y Tecnológica), FONIS Regular SA16I0305 (Lydia Gómez-Pérez, Principal Investigator). This study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Psychometric properties of the Fibromyalgia Survey Questionnaire (FSQ) in Chilean women with fibromyalgia

Abstract

Objective: to evaluate the psychometric properties of the Chilean version of the FSQ. **Methods:** Women with fibromyalgia (FM, *n* = 214), women with rheumatoid arthritis (RA, *n*= 97), and women without chronic pain (attended at the Gynecologist, G, *n* = 117) from the *Red Salud UC-Christus* (Santiago, Chile) participated. Women with FM completed the *Fibromyalgia Survey Questionnaire*; *Fibromyalgia Impact Questionnaire Revised*; *Numerical Pain Rating Scale; Pain Catastrophizing Scale; Pain Vigilance and Awareness Questionnaire; Patient Health Questionnaire-15; and Short-Form Health Survey*. Two weeks later, they completed the FSQ again by phone (n=120). **Results:** the FSQ total scale showed excellent to good internal consistency at T1 (α= .91, ω= .91) and T2 (α= .78, ω= .78) and good test-retest reliability (ICC=.79; 95%CI: .72-.85). It showed medium to large correlations with the other measures. Discriminant analysis between FM group and Control group (RA and G) revealed that the FSQ total scale reached a classification accuracy of 81.3%. ROC curve (AAUC= .88; 95%CI: .85-.92) showed that the best FSQ cutoff was 17, resulting in sensitivity 89% (95%CI:.84-.93) and specificity 75% (95%CI:.69-.80). Considering the FM diagnosis performed by a rheumatologist as the gold standard, sensitivity and specificity of the modified 2010 ACR preliminary criteria for FM were 92.8% (95%CI:.88-.96) and 63.4% (95%CI:.57-.70), respectively. **Conclusion:** the Chilean version of the FSQ presents good psychometric properties and is a useful tool in clinical settings to assist in FM diagnosis and symptomatology assessment. A cutoff score of ≥ 17 appears to be the most appropriated for Chilean population.

Key Indexing Terms: fibromyalgia; illness index severity; musculoskeletal pain; outcome assessment.

Introduction

Fibromyalgia (FM) is a disorder, characterized by chronic, generalized musculoskeletal pain, and the presence of other psychosomatic, cognitive, and neurological symptoms (1-3). It may be difficult to diagnose, given that it is highly comorbid with several rheumatic, neurological, and psychiatric conditions (4-7) and that its symptoms overlap with the symptoms of these conditions (4, 8-9). In 2010, the American College of Rheumatology (ACR) adapted the FM diagnostic criteria (3) by changing the focus from physical pain, to FM's multidimensional nature. Based on these changes, the Fibromyalgia Survey Questionnaire (FSQ) was developed, which has shown to be useful to diagnose FM in epidemiological and clinical studies (10-11). The FSQ consists of two scales: The Widespread Pain Index (WPI) and the Symptom Severity Scale (SS). The WPI consists of a list of 19 non-articular regions in which patients indicate whether they have had pain or tenderness. WPI scores range from 0-19. The SS assesses the extent and intensity in which patients experience six core symptoms: sleep disturbance, fatigue, cognitive problems, headaches, abdominal pain or cramps, and depression. SS scores range from 0 to 12. The sum of WPI and SS constitutes the FSQ total Scale (also referred as the Polysymptomatic Distress Scale, PSD), which not only allows establishing a FM diagnosis (Score 0-31; cut-off score ≥ 13), but also assessing the intensity and symptomatic severity of the disorder (3, 11-13). Thus, FM can be studied on a dimensional or continuum scale, allowing further exploration of the syndrome (8, 14-16). FM diagnosis is made when the following conditions are met: 1) Generalized pain, defined as pain in at least 4 or 5 regions, is present; 2) Symptoms have been present at a similar level for at least 3 months; 3) WPI ≥ 7

and SS ≥ 5 or WPI= 4-6 and SS ≥ 9. Currently, a FM diagnosis is valid irrespective of other diagnoses, and it does not exclude the presence of other illnesses (17); however, we started participants' recruitment before this criterion was stablished and, therefore, participants with other illnesses were excluded in the present research. Conditions 1, 2, and 3 can be assessed by FSQ (11), which works as a screening tool. Nonetheless, in order to confirm the diagnosis, differential diagnosis performed by a specialist is needed. Therefore, the FSQ is not meant to replace or remove the use of other resources or diagnostic criteria, but rather, it can be understood as an alternative method to support early diagnosis and treatment of FM in various clinical contexts, and at a longitudinal level (3,11).

So far, the FSQ has been validated in Canadian (18), French (19), American (20), Japanese (21), Iranian (22-23), Turkish (24), and Spanish populations (25). It has shown acceptable levels of sensitivity (with ranges from 64% to 96%), specificity (ranging from 60% to 100%), reliability (with α ranges between .60 and .85), and validity (significant correlation coefficients with different questionnaires, such as FIQ-R, PHQ, and SF-12).

Despite the fact that the FSQ has shown to be useful to diagnose FM in epidemiological and clinical studies, and have been validated in many countries, no previous studies have examined the psychometric properties of the FSQ in Chile. Furthermore, cut-off scores for Chilean population are not available yet. Therefore, in our study, we aimed to evaluate the psychometric properties of the Chilean version of the FSQ. Namely, we assessed its reliability, (internal consistency and test-retest reliability), and discriminant validity. We also calculated the best FSQ cutoff score for this population and assessed its sensitivity and specificity.

Materials and Methods

Study design and participants

Adult women with FM (n = 214) and women with Rheumatoid Arthritis (RA, n= 97) that attended *Red Salud UC-Christus* (Santiago, Chile) participated in the present study. Additionally, adult women without chronic pain attending the Gynecology Unit of the same institution participated (G, n = 117). Exclusion criteria were being pregnant, presence of cancer, systemic vascular disease, mild to severe neurological problems, inflammatory rheumatic diseases, and comorbidity between FM and RA. A total of 431 subjects were recruited in the study. From this sample, 19 FM were excluded due to exclusion criteria, and 5 (FM = 1, RA= 1, G= 3) were excluded due to FSQ not being completed. As a result, a total sample of 407 (FM= 194, RA= 96, G=117) subjects were used for statistical analysis.

Instruments

Fibromyalgia Survey Questionnaire (FSQ) (11). The FSQ assesses the main FM symptoms according to the 2010 ACR preliminary diagnostic criteria (3). The elaboration of this instrument implied the modification of these preliminary criteria so that the diagnosis can be based on the symptoms self-reported by the patients (11). The Spanish version of the test (25) was adapted to be used in Chilean population.

Fibromyalgia Impact Questionnaire, revised version (FIQ-R) (26). The FIQ-R comprises 21 items in a 0-10 rating scale, which assesses physical functioning and FM symptom severity. It has three subscales (Functioning, Pain Impact, and Symptom Severity). Functioning scale includes nine items, assessing difficulty performing different daily activities. Pain impact scale consists of two items related to FM general impact. Finally, Symptom Severity scale has 10 items involving symptoms commonly reported by FM patients. A total score of the questionnaire can also be computed. The Spanish version of this questionnaire possesses adequate psychometric properties (27). We assessed whether this version of the

questionnaire was considered appropriate by Chilean patients using the sample of the present research (see method section). The Chilean version of this instrument showed adequate construct validity, internal consistence (ω= .96 for the total score) and test-retest reliability (ICC = .90) (28).

Numerical Pain Rating Scale (NPRS) (29). *It* consists of four numeric scales ranging from 0-10, assessing pain intensity at the time of the interview, as well as the worst, the slightest and the average pain experienced during the week.

Pain Catastrophizing Scale (PCS) (30). It is the most frequently used instrument to assess pain-related catastrophic thoughts and emotions. It includes 13 self-report items in a Likert scale (ranging from zero to four), comprising three subscales: Rumination, Magnification, and Hopelessness. A total score can also be calculated. Its Spanish version possesses adequate psychometric properties (31). Furthermore, it has been previously validated in Chile in two samples of pregnant women (32, 33). In the present sample, we assessed whether the previous Chilean version of the questionnaire was considered appropriate by Chilean patients with FM (see method section). Further analyses conducted with this sample indicated that this instrument showed appropriated construct validity, internal consistency (Cronbach alpha was .94 for total PCS score), and test-retest reliability (ICC was .84) (34).

The Pain Vigilance and Awareness Questionnaire (PVAQ) (35). This instrument evaluates pain hypervigilance. It consists of nine items rated from zero to five and organized in two subscales: Active Vigilance and Passive Awareness. The psychometric properties of this scale are adequate. A total score of the questionnaire can also be calculated. We also assessed whether the Spanish version of this questionnaire was considered appropriate by Chilean patients using the present sample (see method section). Preliminary analyses of the Chilean

version of this instrument showed acceptable construct validity, internal consistency (Cronbach alpha was .78 for the total score), and test retest reliability (ICC = .83) (36).

Patient Health Questionnaire-15 (PHQ-15) (37). The PHQ-15 is a somatic symptom subscale that assesses the presence of physical problems during the last 4 weeks. It comprises 15 items, rated from zero-two. Patients can be classified in four categories according to symptom severity: (scores = 0–4), low (scores = 5–9), medium (scores = 10–14), and high (scores = 15–30). The Spanish version was used (38), which has adequate psychometric properties. We also assessed whether the Spanish version of this questionnaire was considered appropriate by Chilean patients using the present sample (see method section). In the present sample, the internal consistency of the PHQ-15 was ω= .78.

Short-Form Health Survey (SF-12) (39). It is a quality of life questionnaire which appraises degree of well-being and functional ability. It is a shorter version of the SF-36. It consists of 12 items related to eight dimensions (Physical Function, Social Function, Physical Role, Emotional Role, Mental Health, Vitality, Bodily Pain, and General Health). We used the Chilean version of this questionnaire, which possess good psychometric properties (40). In the present sample, the internal consistency of the SF-12 was ω= .86.

Brief Pain Inventory Interference subscale (BPI-IS) (41). It assesses the degree of pain interference in daily life. It comprises seven items rated from zero to ten, assessing different areas: General Activity, Mood, Walking Ability, Normal Work, Relationships with other people, Sleep, and Enjoyment of Life. In the present study, we used the Spanish version of this questionnaire, which have shown adequate reliability and validity (42). Whether the Spanish version of this questionnaire was considered appropriate by Chilean patients was also

examined using the present sample (see method section). In the present sample, the internal consistency of this subscale was ω= .91.

Patient Health Questionnaire-9 (PHQ-9) (43). It assesses the presence of depressive symptoms in the last 2 weeks, according to the DSM-IV criteria. It consists of nine items, rating level of severity (zero to three). According to the obtained score and referred symptoms, diagnostic categories can be identified, as well as symptomatology type and severity, including: Major Depressive Syndrome, Other Depressive Syndrome, Positive Depressive Symptoms, and Negative Depressive Symptoms. The Chilean validated version was used in this study (44). In the present sample, the internal consistency of the PHQ-9 was ω= .84.

Procedure

This study was approved by the Medicine Scientific Ethics Committee from the Pontificia Universidad Católica de Chile (protocol number 16-210). In order to assess whether the Spanish version of the questionnaire was considered appropriate by Chilean patients, before collecting the study sample, this version of the scale, as well as other instruments not yet validated in Chile at the moment of the study in patients with FM (namely the PCS, FIQ-R, the PVAQ, the PHQ-15, and the BPI-IS) were applied to ten Chilean women with FM that were attending the rheumatology unit of the Centro Médico San Joaquín, which is part of Red Salud UC Christus. They were asked to indicate whether the items of these questionnaires were comprehensible, and the wording was clear to Chilean speakers. They could openly suggest any changes in the wording they deemed necessary to improve the scale. All items were considered comprehensible for all the participants. Similarly, they indicated that the items were worded in a way that they were adequate for its use in Chilean population, and no

changes were suggested. We did not consider these data in the statistical analyses conducted in order to assess the psychometric properties of the FSQ.

Participants with FM and participants with RA were invited to participate in the study by a research assistant when they were waiting to be attended by their rheumatologist. Additionally, FM cases were identified using a computer-based system that sent weekly notices of subjects with a diagnosis of fibromyalgia in electronic clinical notes. Furthermore, flyers and banners were placed in the rheumatology clinic waiting room with contact information when a research assistant was not present. We confirmed participants diagnoses by asking participants' physicians. Furthermore, we specifically asked the physicians to double-check that those patients with FM did not present RA, and vice versa. If the doctor was not available to confirm her patients diagnoses, we verified it using clinical records. Women without chronic pain were invited to participate while they were waiting to be attended by their gynecologist. The research assistant explained the purpose of the study and obtained written informed consent to participate as well as to publish the results of the study. Then, she asked participants to complete the questionnaires. Participants with FM completed the full battery of questionnaires, while participants with RA and participants without chronic pain only completed the FSQ, the NPRS, and some questions regarding clinical and sociodemographic characteristics. When a participant reported having vision or reading problems that prevented them to fill the questionnaires, they received help from the research assistant to complete them. In order to examine test-retest reliability of the FSQ, we asked participants with FM to complete this questionnaire again, this time by phone, two weeks later.

Statistical analysis

Psychometric properties of the FSQ were assessed using R-Statistical version 3.5.1 and SPSS version 24. Before conducting the main analyses, we examined whether there were differences between the groups regarding the sociodemographic variables (age, region, marital status, educational level, and employment status). We conducted a non-parametric analysis (Kruskal Wallis Test) to check for differences in age, because the groups differed in sample size; and Chi-squared Tests to examine differences between the groups in the categorical variables. Then, we performed the following analyses:

Reliability

Internal consistency. We computed Cronbach's alpha (α) and McDonald's omega coefficient (ω) for the SS and the FSQ total scales at T1 and T2, using all the participants of the study. We considered scores above .70 for alpha and omega as satisfactory (45).

Test–Retest reliability. We calculated the intra-class correlation coefficients (ICCs) between T1 and T2 of the SS, the WPI, and the FSQ total scales, using only the sample of participants with fibromyalgia. We used a 95% confidence interval, based on a single-rating, absolute-agreement, two-way mixed-effects model. We used the following ICC values range: poor (<.50), moderate (.50-.75), good (.75-.90), and excellent agreement (>.90) (46). For this data analysis, only subjects who participated in the follow-up [n=120, 61.9%] were considered.

Validity

In order to assess convergent validity, we calculated Pearson's correlations between the FSQ total scale in FM group and the additional instruments (FIQ-R, NPRS, PCS, PVAQ, PHQ-15, SF-15, BPI, and PHQ-9). In order to assess discriminant validity, we conducted two hierarchical discriminant analyses. The first one considering the score in the FSQ total scale as the

independent variable and predicting whether participants belonged to the FM group or the Control group (RA + G). The second one considering only the FM and the AR groups and examining whether the FSQ could discriminate between these samples. In both analyses, we introduced age, educational level, and marital and employment status as control variables in the first step of the model. Then, we introduced the FSQ total score in the second step and examined whether the accuracy of the model was improved by the inclusion of this variable, once the covariables had been controlled for. Before conducting the analyses, categorical variables with more than two categories were transformed into dummy variables. For marital status the baseline category was "Married/Live-in partner"; for educational level the baseline category was "high-school education or less", and for employment status the baseline category was "being working" (see Table 1). Following Huberty's suggestions (47) we calculated the discriminant function with half of the sample, and then cross-validated these results with the other half of the sample. The levels of accuracy reported are the ones calculated during the cross-validation phase.

Cut-off score, sensitivity, and specificity

We conducted receiver operating characteristic curve (ROC, 48) in order to determine the Area Under the Curve (AUC) and the best FSQ total scale cut-off score. In order to calculate the FSQ total scale Adjusted AUC (AAUC), we conducted Covariate-Adjusted ROC (49) controlling for age, educational level, and marital and employment status. Finally, we examined the sensitivity and specificity for the FSQ modified 2010 preliminary criteria for FM recommended by Wolfe et al. (11), that is presenting a WPI score ≥ 7 and a SS score ≥ 5 or presenting a WPI score equal to 3-6 and a SS score ≥ 9 , as well as presenting these symptoms for at least three months.

Sample size calculation. The sample size needed for the abovementioned analyses was calculated a priori. First, we used the program EPIDAT 4.2. (50) to determine the sample needed to calculate sensitivity and specificity. For that, we used the values reported by Wolfe et al (11), that is 96,6% of sensitivity and 91.8% of specificity, with a ratio of sick vs no sick of 1, an absolute precision of 5% and a confidence interval of 95%. The results of these analyses indicated that we needed 233 participants in total (117 in the fibromyalgia group and 116 in the non-fibromyalgia group). Then we used the software GPower 3.1. (51) to calculate the sample size needed for the discriminant analyses, using the following parameters: a medium size effect of f2V =.0625, with an alpha level = .05, and power = .80. The results indicated that we needed a total sample of 128. Then, we used the program MedCalc (52) to calculate how many participants we needed to conduct the Roc curve. The result of these analyses showed that for an alpha of .01 and a beta (probability of type two error) of .01, with an area under the curve of .70, a null hypothesis value of .05, and a ratio of sample size in negative/positive groups equal to 1, we required 182 participants in total, 91 in each group. The specificity and sensitivity analyses were the ones requiring a higher sample size (233 in total). Nonetheless, we recruited a larger sample of participants (the total sample size used for statistical analyses was 407, FM= 194, RA= 96, G=117). The reason was that we were simultaneously recruiting participants with research purposes other than the ones reported in the present manuscript. In this way we a priori assure we had the power we needed to perform our statistical analyses.

Results

We present the demographics and clinical characteristics of the sample in Table 1 and 2. The descriptive statistics of each instruments for each group are described in Table 3. The sample was mainly composed of middle age women, from Chile's capital Santiago (84.3%),

predominantly married/live-in partner (58%), with a full/part time employment (50.9%) and a high school education level or less (36.6%).

Significant differences were found between the groups in the mean age of the participants, RA being the oldest group and G the youngest one. The groups also significantly differed in marital status, educational level, and employment status (See Table 1).

Participants with FM presented an average of 7.46 years with pain and 3.61 years since their diagnosis. They had a significantly higher score in the FSQ total scale (F[2, 404] =181.99, p <.0001), Symptom Severity Scale (SS) (F[2, 404] =86.01, p <.0001), Widespread Pain Scale (F[2, 404]=177.86, p <.0001) and consumption of antidepressants (92.5%) than other groups (7.5%) (See Table 2). The most frequently used antidepressant type was serotonin and norepinephrine reuptake inhibitors (SNRIs, 67.9%).

Reliability.

Internal consistency reliability. FSQ total scale indicates excellent to good internal consistency at T1 (α= .91, ω= .91) and T2 (α= .78, ω= .78). However, SS only shows good internal consistency at T1 (α= .74, ω= .76) but a questionable internal consistency at T2 (α= .60, ω= .62).

Test-Retest reliability. The SS showed moderate test-retest reliability (ICC = .65 [.53- .74]), while the WPI (ICC= .75 [.66-.82]) and PSD (ICC=.79 [.72-.85]) showed good test-retest reliability.

Validity.

Convergent validity. Correlations between PSD and other measures were mostly medium to large and statistically significant (See Table 3, SDC 3, descriptive statistics of the instruments used). There were large and positive correlations with the PHQ-15 (r=.62, p<.0001), the FIQ-

R (r=.60, p<.0001), and the NPRS (r=.51, p<.0001). The PSD showed medium and positive correlations with the PHQ-9 (r=.49, p<.0001), the BPI-PI (r=.47, p<.0001), and the PCS (r=.31, p<.0001); and an inverse and medium association with the SF-12 (r= -.46, p<.0001). Finally, related to PVAQ, it presented a small and positive correlation (r=.22, p=.002).

Discriminant validity. The results of the hierarchical discriminant analysis conducted to predict whether participants belong to the FM or Control group (RA+G) showed that there were significant mean differences between the groups in the predictors 'FSQ total score' (Wilk's λ =.612, p<.0001) and 'being widowed' (Wilk's λ =.963, p=.006). In the first step, when only the control variables were included in the analyses, the linear equation significantly predicted group membership (Wilk's λ = .883, 2 (11) = 24.36, p=.011), accounting for 11.6% of between-group variability. The relative contribution of each of the control variables is reported in the structure matrix presented in Table 4. A loading higher than .30 is considered as the cut-off between important and less important variables (53). The variable with the higher contribution to the equation was being married/live-in partner versus being widowed, followed by being working versus being a student, being working versus being retired, being working versus being in medical leave, having high-school education or less versus having technical studies, and being married/live-in partner versus being single. Namely, being married/live-in partner (when compared with those widowed and single), being a student (instead of a worker), and being working (instead of retired) were variables inversely associated with having FM; whereas being working (versus in medical leave) and having a high-school education or less were directly associated with having FM. The classification accuracy was 56.7%. When the FSQ total score was introduced in the second step, the model significantly improved (Wilk's λ = .561, 2 (12) = 113.37, p <.0001)

318 accounting for a higher between-group variability (44%) and showing a classification
319 accuracy of 81.3%. The relative contribution of each of the control variables is reported in the
320 structure matrix presented in Table 4. Only the FSQ total score significantly contributed to
321 the linear equation (*r* =.90).
322 The results of the hierarchical discriminant analysis conducted to predict whether
323 participants belonged to the FM or the AR group showed that there were significant
324 differences between the group in the following predictors: mean FSQ total score (Wilk's
325 λ=.721, *p*<.0001), being married/living-in partner versus being widowed (Wilk's λ=.923,
326 *p*=.001), being working versus retired (Wilk's λ=.952, p=.008), age (Wilk's λ=.955, p=.010),
327 having high-school education or less versus technical school education (Wilk's λ=.961,
328 *p*=.018), and being married/living-in partner versus being single (Wilk's λ=.973, *p*=.048). In
329 the first step, when only the control variables were included in the analyses, the linear
330 equation significantly predicted group membership (Wilk's λ = .795, 2 (11) = 31.59, *p*
331 =.001), accounting for 20.5% of the between group variability. The relative contribution of
332 each of the control variables is reported in the structure matrix presented in Table 4. The
333 predictors that presented a cutoff score higher than .30 were: being married/live-in partner
334 versus being widowed, followed by being working versus retired, age, having high-school
335 education or less versus technical school education, and being married/living-in partner
336 versus being single, respectively. Specifically, the age, being married (when compared with
337 those widowed or retired), being working (instead of retired) was inversely associated with
338 being a member of the FM group; whereas having high-school education (versus technical
339 school education) was directly associated with being a member of the FM group. The
340 classification accuracy was 62.1%. When, the FSQ total score was introduced in the second

step, the model significantly improved (Wilk's λ = .573, 2 (12) =76.37, p<.0001) accounting for a higher between group variability (42.8%) and showing a significantly higher classification accuracy of 71.7%. The relative contribution of each of the control variables is reported in the structure matrix presented in Table 4. Only the FSQ total score significantly contributed to the linear equation (r =.72).

Cut-off score, sensitivity, and reliability. ROC curves showed that the FSQ total scale as a predictor for the classification to the groups (FM and control [RA + G]) had a very good performance to discriminate the presence versus absence of fibromyalgia (AUC= .88; 95%CI: .84-.91), and a Covariate Adjusted ROC showed an almost identical result, (AAUC= .88 :95%CI: .85-.92) (see Figure 1). The best cutoff that maximizes (sensitivity + specificity) is 17 (see Table 5), resulting in sensitivity .89 (95% CI: .84-.93), specificity .75 (95% CI: .69-.80), Positive Predictive Value (PPV) .76, Negative Predictive Value (NPV) .88, Positive Likelihood Ratio (PLR), 3.50, and Negative Likelihood Ratio (NLR) .15.

When using the modified 2010 ACR preliminary criteria for fibromyalgia (11) to classify the patients, instead of the FSQ total score, the level of sensitivity was high (92.8%; 95% CI:.88-.96) and the level of specificity was adequate (63.4%; 95% CI: .57-.70). Nonetheless, FSQ specificity was higher in G group (77.8%; 95% CI: .69-.85) than the RA group (45.8%; 95% CI: .36-.56).

Discussion

The present study showed acceptable reliability and validity for the FSQ as a fibromyalgia diagnostic instrument. PSD (total scale) and WPI presented good performance, especially regarding test-retest reliability, which indicated consistent answers through time. In addition, in line with our expectations, participants with FM presented higher SS, WPI and FSQ total

scale scores than participants in the control groups. Furthermore, the FSQ showed good convergent validity, as it presented positive and large correlations with the other measures. Mainly, the PSD was associated with somatic symptom severity, fibromyalgia impact, and pain intensity. Similar results were observed in previous validations (e.g. 19, 25), which provides more evidence of the impairment that physical and psychosocial symptoms of FM elicit on patients.

Despite this, it is important to consider the difference of results between the scales, as SS did not perform as well as the others scales, decreasing in .14 its internal consistency at T2 [α= .60, ω= .62] and getting a moderate test-retest reliability [ICC .62]. One explanation may be that some of the participants were first interviewed after receiving their first consultation for FM with a rheumatologist, and then interviewed again two weeks after starting a new medical treatment. Therefore, symptoms' improvement may be responsible for the lower internal consistency at T2 and the moderate test-retest reliability. Furthermore, the first rheumatologist assessment itself (and not only the medical treatment) can produce an initial effect in the symptoms, which could be affecting the stability of the SS scores. An alternative explanation is that nine of the 12 scores of the SS explore symptoms experienced during the last week, which usually fluctuate overtime. Although most measures show stability over time, SS variations at test-retest must be understood considering FM's dynamic nature.

When compared with our gold standard (rheumatologist diagnosis), the modified 2010 ACR preliminary criteria for FM (11) (WPI ≥ 7 and SS ≥ 5 or WPI 3–6 and SS ≥ 9) showed greater sensitivity (92.8) but lower specificity (63.4) than the FSQ total score (.89 and .75 respectively). The high sensitivity found in the modified 2010 ACR preliminary criteria and the FSQ total score shows that this instrument can correctly identify FM subjects. The FSQ

total score presented better specificity properties, and a better balance between sensitivity and specificity, than the modified 2010 ACR preliminary criteria (11).

The best cut-off score that maximizes sensitivity and specificity properties of the FSQ was 17 for the Chilean version, which is higher than the recommended by Wolfe et al. (11). Considering that patients with FM are a heterogeneous group, a possible explanation for the different cutoff scores can be cross-cultural differences, a trend that has also been reported by other studies (e.g. 21-22). It must also be considered that some longitudinal studies of patients with FM have shown fluctuations around the cutoff points (54-55). The presence of FM subgroups or clusters have been suggested, based on variations in pain and other symptomatology (14, 56), that might cause significant differences in the way patients experience and report the syndrome, especially in terms of psychosocial factors. As FM is still a highly complex condition, and diagnosis and assessment criteria keep being developed, FSQ has some limitations as well. For example, the fact that it is almost completely reliant on symptoms, and doesn't include the assessment of mechanistic factors (57-58), which could be a necessary change, as central sensitization is a core pathophysiological factor in FM diagnosis and assessment (2, 59-60).

FSQ is a fast self-administered, easy to understand questionnaire. Some advantages of our study are that this is the first validation of FSQ in Latin America. This is relevant given cultural differences may affect pain experiences. In addition, we studied an adequate sample size, controlled for differences in sociodemographic variables, and were able to follow up subjects, which is a critical aspect when assessing FM, as it is a dynamic condition with significant changes in symptomatology and presentation over time.

This study, however, presents some limitations. First, diagnosis was provided by the physicians attending the participants, and the research team did not conduct any standard healthcare professional examination as part of the research procedure. Nonetheless, we specifically asked the physicians to confirm that the patients with FM did not present RA, and vice versa. In spite of this, fibromyalgia could eventually be present among patients with RA, especially if we take into account that until recently, the diagnosis of FM was ruled out when the participant presented another inflammatory disease that could be otherwise explaining the pain (11). This could have impacted our cut-off score for fibromyalgia diagnosis and be partially explaining why our cut off score (≥17) was higher than the one previously reported in other countries (≥13). This potential co-ocurrence of FM and RA may have also affected the results of our discriminant analyses. Finally, this could also at least partially explain the low specificity found for the FSQ modified ACR 2010 preliminary diagnosis criteria recommended by Wolfe et al. (11) when using the participants with RA as the comparison group. Replicating our results in future studies in which this comorbidity is more rigorously controlled as well as in studies comparing the specificity and sensitivity of the FSQ using other chronic pain conditions as the comparative group is needed. Despite this limitation, this is the first study in which the validity of the FSQ is examined separating RA in a single group. In previous studies, the comparison non-fibromyalgia group generally comprises RA patients as well as patients with other pain disorders such as osteoarthritis (OA). However, the proportion of patients with noninflammatory chronic pain diagnoses such as OA (which frequently affects only a few joints and does not produce fatigue) is usually much higher than the proportion of patients with RA. Therefore, it is possible that in future studies using a comparison group comprising only patients with RA, the cutoff scores could be higher than the reported until

now in the literature. In this sense, our findings contribute to the literature by providing for the first time data regarding the validity of the FSQ to discriminate between patients with FM and patients with RA, and suggest that the ability of the test to discriminate between FM and RA may be lower than the ability to discriminate FM amongst other chronic pain diagnoses. Second, we considered the FM diagnosis performed by a physician as the gold standard and, therefore, our findings are subject to the same limitations and controversies that nowadays surround the FM diagnosis. In this sense, we agree with Galvez-Sánchez & Reyes del Paso (61) in that the current lack of clear consensus regarding the concept and diagnosis of FMS amongst medical professionals remains, and that complaints from health professionals and patients about the way the disease is diagnosed continue. Third, usually in studies analyzing test-retest reliability, not all the participants, but a subsample is contacted. As such, we did not invite all the study participants to complete the second assessment aimed to assess the FSQ test-retest reliability. Nonetheless, one limitation of the present research is that we did not adequately register how many participants were finally contacted for the test-retest and, therefore, we do not know which was the dropout rate for the second assessment. In case that the dropout was higher than 20%, the external validity of our results would have been compromised. However, although we do not have reasons to think that the dropout rate of our study was higher than the one presented in previous research, our results regarding the test-retest reliability of the FSQ must be taken with caution. In spite of this constraint, only two previous studies have examined the test-retest reliability of the FSQ (19, 24), and unlike our study, both of them have the limitation of having used Pearson´s correlations – instead of intraclass correlation – to assess this type of reliability, which is not appropriate (62). Therefore, despite our limitations, our study provides results about the reliability of the FSQ

that are not available in the current literature and overcomes the limitations of these two previous studies. Conducting more rigorous studies assessing the retest-retest reliability of the FSQ is in the future is indeed essential. As FM symptom variations are expected, it is important to have strong longitudinal reports to be able to capture the dynamics of the pain experience. Another limitation of the present study is that two of the questionnaires used to assess the FSQ convergent validity (i. e. the PHQ-15 and the interference subscale of the BPI) had been previously validated in Spain, but not in Chile. Furthermore, although the construct validity and reliability of two of the scales (the FIQ-R and the PVAQ) have been previously examined using the same sample as in the present research, the convergent validity of these two instruments remains to be conducted in Chile. Although these four instruments showed adequate reliability in the present sample, the results regarding the convergent validity using these questionnaires should be taken with caution and further studies about the convergent validity of these questionnaires need to be conducted. In addition, as the majority of the studies conducted about the validity of FSQ, participants were not community subjects, which might have limited the generalizability of our results. Moreover, we did not include patients from other settings, such as primary care. Furthermore, we recruited only female patients, so the results cannot be generalized to men with FM. Besides, discriminant analyses were performed using patients with RA and participants without chronic pain and, did not include patients with other diagnoses (such as osteoarthritis, spondyloarthropathies, polyneuropathy, etc.); therefore, the discriminant value of the test is limited only for patients with RA. Studies including participants with other pain diagnoses should be conducted in the future. It is also important to consider variations in T1 and T2 assessments, as they had

different conditions (e.g. T1 involved a face-to-face interview, while T2 was conducted over the phone).

In conclusion, the Chilean version of the FSQ has shown good psychometric properties. It is a useful tool in clinical settings to assist in FM diagnosis and symptomatology assessment. Setting a cutoff score of ≥ 17 appears to be the most effective approach in Chilean population.

Acknowledgment

We want to thank to all the women that participated in the present research as well as the reviewer of the paper for their time and insights. We also want to thanks to the research assistants Camila Belén Román Butrón, Débora Martellans, Alejandra Escobar Zuluaga, Ana Carolina de Obaldía, Alejandra Rivera, Claudia Toloza, Francisca Navarro, Javiera Soza, Laura Maldonado, Xaviera Acuña, Manuela Pacheco, and Montserrat Fellenberg, who contributed to the data collection. We also want to thank Dr. Jorge Flores, who helped edit the final version of the manuscript.

References

1. Arnold L, Clauw D, McCarberg B. Improving the Recognition and Diagnosis of Fibromyalgia. *Mayo Clin Proc.* 2011; 86:457-464.
2. Sarzi-Puttini P, Atzeni F, Masala I, et al. Are the ACR 2010 Diagnostic Criteria for Fibromyalgia Better than the 1990 Criteria? *Autoimmun Rev*. 2018; 17:33-35.
3. Wolfe F, Clauw D, Fitzcharles M, et al. The American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and Measurement of Symptom Severity. *Arthritis Care & Res*. 2010; 62:600-610.
4. Arnold L. The Pathophysiology, Diagnosis and Treatment of Fibromyalgia. *Psychiatr Clin North Am.* 2010; 33:375-408.
5. Bennett R. Diagnostic Criteria and Differential Diagnosis of the Fibromyalgia Syndrome. *J Musculoskelet Pain.* 2004; 12:59-64.
6. Natelson B. Fibromyalgia Syndrome: Diagnosis and Comorbidities. *CNS Spectr*. 2009; 16:8-11.
7. Russell I, Raphael K. Fibromyalgia Syndrome: Presentation, Diagnosis, Differential Diagnosis, and Vulnerability. *CNS Spectr.* 2008; 15:6-11.
8. Häuser W, Sarzi-Puttini P, Fitzcharles M. Fibromyalgia Syndrome: Under-, Over, and Misdiagnosis. *Clin Exp Rheumatol*. 2019; 116:90-97.
9. Valls-Llobet C. Diagnóstico Diferencial del Dolor y de la Fibromialgia. *Anu Psicología.* 2008; 39:87-92.
10. Häuser W, Wolfe F. Diagnosis and Diagnostic tests for Fibromyalgia (syndrome). *Reumatismo*. 2012; 64:194-205.
11. Wolfe F, Clauw D, Fitzcharles M, 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; 38:1113-1122.

12. Friend R, Bennett R. A Critical Examination of the Polysymptomatic Distress Scale Construct as a Symptom Severity Questionnaire. *J Rheumatol.* 2015; 42:1364-1367.
13. Wolfe F, Walitt B, Rasker J, et al. The Use of Polysymptomatic Distress Categories in the Evaluation of Fibromyalgia (FM) and FM Severity. *J Rheumatol*. 2015; 42:1-8.
14. Bidari A, Ghavidel-Parsa B, Ghalehbaghi B. Challenges in Fibromyalgia Diagnosis: From Meaning of Symptoms to Fibromyalgia Labeling. *Korean J Pain.* 2018; 31:147-154.
15. Wang S, Han C, Lee S, at al. Fibromyalgia Diagnosis: A Review of the Past, Present and Future. *Expert Rev Neurother.* 2015; 15:667-669.
16. Wolfe F, Brähler E, Hinz A, et al. Fibromyalgia Prevalence, Somatic Symptom Reporting, and the Dimensionality o Polysymptomatic Distress: Results from a Survey of the General Population. *Arthritis Care Res*. 2013; 65:777-785.
17. Wolfe F, Clauw D, Fitzcharles M, et al. Revisions to the 2010/2011 Fibromyalgia Diagnostic Criteria. *Semin Arthritis Rheum*. 2016; 46:319-329.
18. Ferrari R, Russell A. A Questionnaire Using the Modified 2010 American College of Rheumatology Criteria for Fibromyalgia: Specificity and Sensitivity in Clinical Practice. *J Rheumatol*. 2013; 40:1590-1595.
19. Fitzcharles M, Ste-Marie P, 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; 13:1-6.

20. Bennet R, Friend R, Marcus D, et al. Criteria for the Diagnosis of Fibromyalgia: Validation of the Modified 2010 Preliminary American College of Rheumatology Criteria and the Development of Alternative Criteria. *Arthritis Care Res.* 2014; 66:1364-1373.
21. 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* 2012; 23:846-850.
22. 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.
23. Bidari A, Ghavidel-Parsa B, Amir A, et al. Validation of Fibromyalgia Survey Questionnaire and Polysymptomatic Distress Scale in a Persian Population. *Rheumatol Int.* 2015; 35:2013-2019.
24. Yanmaz M, Atar S, Biçer M. The Reliability and Validity of the Turkish Version of Fibromyalgia Survey Diagnostic Criteria and Symptom Severity Scale. *J Back Musculoskel Rehabil.* 2016; 29:287-293.
25. Carrillo de la Peña M, Triñanes Y, González-Villar A, et al. Convergence Between the 1990 and 2010 ACR Diagnostic Criteria and Validation of the Spanish version of the Fibromyalgia Survey Questionnaire (FSQ). *Rheumatol Int.* 2014; 35:141-151.
26. Bennett R, Friend R, Jones K, et al. The Revised Fibromyalgia Impact Questionnaire (FIQR): Validation and Psychometric Properties. *Arthritis Res Ther.* 2009; 11:415.
27. Salgueiro M, García-Leiva J, Ballesteros J, et al. Validation of a Spanish Version of the Revised Fibromyalgia Impact Questionnaire (FIQR). *Health Qual Life Outcomes.* 2013;

1:132.

28. Escobar A, Román Butrón C, Vergés A, et al. *Psychometric properties of the Chilean version of the Fibromyalgia Impact Questionnaire Revised (FIQ-R) in a sample with Fibromyalgia: Preliminary results* [Conference presentation]. 15th International Congress of Behavioral Medicine (ICBM). Santiago, Chile. 2018, November 14-17.
29. Jensen, M, Turner, J, Romano, et al. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999; 83:157-162.
30. Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and Validation. *Psychol Asses.* 1995; 7:524-532.
31. García-Campayo J, Rodero B, Alda M, Sobradiel N, Montero J, Moreno S. Validación de la Versión Española de la Escala de Catastrofismo ante el Dolor (Pain Catastrophizing Scale) en la Fibromialgia. *Med Clin* 2008; 131:487-492.
32. Martínez M, Miró E, Sánchez A, et al. Spanish Version of the Pain Vigilance and Awareness Questionnaire: Psychometric Properties in a Sample of Women with Fibromyalgia. *Span J Psychol.* 2015; 17:1-8.
33. Esparza C, Soto M, Cumsille P, et al. *Validación de la Escala de catastrofismo ante el dolor en mujeres gestantes* [Conference presentation]. XIV Congreso Chileno de Psicología. Arica, Chile. 2019, November 13-15.
34. Vázquez-Taboada AR, Gómez-Pérez L, Torres-Flores MP et al. *Propiedades psicométricas de la Escala de Catastrofismo ante el Dolor (PCS) en una muestra de mujeres chilenas embarazadas* [Conference presentation]. Symposium Prediciendo la recuperación postparto at the VII Congreso Latinoamericano de Psicología de la Salud

591 [ALAPSA]. Cali, Columbia.2016, May 23-27.
592 35. Gómez-Pérez L, de Obaldía AC, Escobar A, et al. *Psychometric properties of the Chilean*
593 *version of the Pain Catastrophizing Scale in a sample of Chilean women with*
594 *fibromyalgia* [Conference presentation]. 15th International Congress of Behavioral
595 Medicine (ICBM). Santiago, Chile.2018, November 14-17.
596 36. Román C, Escobar A, Arrianza T, et al. *Psychometric properties of the Chilean version*
597 *of the Pain Vigilance and Awareness Questionnaire (PAVQ) in women with fibromyalgia*
598 [Conference presentation]. 15th International Congress of Behavioral Medicine
599 (ICBM). Santiago, Chile. 2018, November 14-17.
600 37. Kroenke K, Spitzer R, Williams J. The PHQ-15: Validity of a new Measure for
601 Evaluating the Severity of Somatic Symptoms. *Psychosom Med*. 2002; 64:258-266.
602 38. Ros S, Comas A, García-García M. Validación de la Versión Española del Cuestionario
603 PHQ-15 para la Evaluación de Síntomas Físicos en Pacientes con Trastornos de
604 Depresión y/o Ansiedad: Estudio DEPRE-SOMA. *Actas Esp Psiquiatr.* 2010; 38:345-
605 357.
606 39. Ware J, Kosinski, M, Keller S. A 12-item short-form Health Survey. Construction of
607 Scales and Preliminary Tests of Reliability and Validity. *Med Care*. 1996; 34:220-233.
608 40. Vera-Villarroel P, Silva J, Celis-Atenas K, et al. Evaluación del Cuestionario SF-12:
609 Verificación de la Utilidad de la Escala de Salud Mental. *Rev Med Chile.* 2014;
610 142:1275-1283.
611 41. Cleeland C, Ryan K. Pain Assessment: Global Use of the Brief Pain Inventory. *Ann Acad*
612 *Med Singapore*. 1994; 23:129-138.
613 42. De Andrés J, Cruces L, Canos M, et al. Validation of the Short Form of the Brief Pain

Inventory (BPI-SF) in Spanish Patients with Non-Cancer-Related Pain. *Pain Pract.* 2015; 15:643-653.

43. Kroenke K, Spitzer R, Williams J. The PHQ-9 Validity of a Brief Depression Severity Measure. *J Gen Intern Med.* 2001; 16:606-613.
44. Baader M, Molina J, Venezian S, et al. Validación y Utilidad de la Encuesta PH1-9 (Patient Health Questionnaire) en el Diagnóstico de Depresión en Pacientes Usuarios de Atención Primaria en Chile. *Rev Chil Neuro-Psiquiatr*. 2012; 50:10-22.
45. Nunnally J, Bernstein I. *Psychometric theory* (3rd ed.). New York: McGraw-Hill; 1994.
46. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016; 15:155-163.
47. Huberty CJ. *Applied Discriminant Analysis (1st ed.).* Michigan: Wiley-Interscience; 1994.
48. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* 2011; 12: 77.
49. Rodríguez-Álvarez MX, Vanda I. "*ROCnReg: An R Package for Receiver Operating Characteristic Curve Inference with and without Covariate Information*." arXiv2020: 2003.13111v2.
50. Consellería de Sanidade, Xunta de Galicia, España; Organización Panamericana de la salud (OPS-OMS); Universidad CES, Colombia. *Epidat: programa para análisis epidemiológico de datos. Versión 4.2.* 2016. Retrieved from https://www.sergas.es/Saude-publica/EPIDAT?idioma=es
51. Faul F, Erdfelder E, Buchner A, & Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009; 41: 1149-1160.

637 52. Schoonjans F. *Medcalc statistics for biomedical research: software manual*. Belgium:
638 MedCalc;1998.
639 53. Garson GD. *Factor analysis* (Statistical Associates Blue Book Series). Asheboro:
640 Statistical Associates Publishers; 2013.
641 54. Adams E, McElroy H, Udall M, et al. Progression of Fibromyalgia: Results from a 2-year
642 observational Fibromyalgia and Chronic Pain Study in the US. *J Pain Res*. 2016; 9:325-
643 336.
644 55. Wallit B, Fitzcharles M, Hassett A, et al. The Longitudinal Outcome of Fibromyalgia: A
645 Study of 1555 Patients. *J Rheumatol.* 2011; 38:2238-2246.
646 56. Bartley E, Robinson M, Staud R. Pain and Fatigue Variability Patterns Distinguish
647 Subgroups of Fibromyalgia Patients. *J Pain*. 2018; 19:372-381.
648 57. Egloff R, von Känel V, Müller U, et al. Implications of Proposed Fibromyalgia Criteria
649 Across Other Functional Pain Syndromes. *Scand J Rheumatol*. 2015; 44:416-424.
650 58. Wolfe F, Egloff N, Häuser W. Widespread Pain and Low Widespread Pain Index Scores
651 Among Fibromyalgia-positive Cases Assessed with the 2010/2011 Fibromyalgia
652 Criteria. *J Rheumatol.* 2016; 43:1743-1748.
653 59. Wolfe F, Butler S, Fitzcharles M, et al. Revised Chronic Widespread Pain Criteria:
654 Development from and Integration with Fibromyalgia Criteria. *Scand J Pain*. 2019;
655 18:77-86.
656 60. Ghavidel-Parsa B, Bidari A, Hajiabbasi A, et al. Fibromyalgia Diagnostic Model Derived
657 from Combination of American College of Rheumatology 1990 and 2011 Criteria.
658 *Korean J Pain.* 2019; 32:120-128.
659 61. Galvez-Sánchez CM, Reyes del Paso, GA. (2020). Diagnostic Criteria for Fibromyalgia:

Critical Review and Future Perspectives. *J Clin Med.* 2020;9: 1219.

62. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between two Methods of Clinical Measurement. *Lancet.* 1986; 1: 307-310.

Figure 1. Receiver operating characteristic curve (ROC) for the FSQ total scale as a predictor for the classification to the groups (FM and control [RA + G]). Area Under the Curve (AUC) = .88; 95%CI: .84-.91). Covariate Adjusted ROC (AAUC= .88 :95%CI: .85-.92). FM = Fibromyalgia; RA = Arthritis Rheumatoid; and G = Attended at the Gynecologist.

Table 1. *Demographic characterization.*

	FM Group n=194	RA Group n=96	G Group n=117	Total, N= 407	Statistics
Age, Mean (SD)	50.38 (12.02)	56.51 (14.57)	41.64 (14.18)	49.31 (14.33)	H(2) = 56.59, p<.0001
Region, n (%)					
Metropolitana de Santiago (Capital)	169 (87.1)	75 (78.9)	99 (84.6)	343 (84.3)	$\chi^2(2) = 3.25$, p= .19
Others	25 (12.9)	20 (20.8)	18 (15.4)	63 (15.5)	
Marital status, n (%)					
Single	36 (18.6)	23 (24.0)	43 (36.8)	102 (25.1)	$\chi^2(6) = 51.55$, p<.0001
Married/Live-in partner	131 (67.5)	42 (45.8)	63 (53.8)	236 (58.0)	
Divorced	22 (11.3)	8 (8.3)	8 (6.8)	38 (9.3)	
Widowed	5 (2.6)	19 (19.8)	3 (2.6)	27 (6.6)	
Educational level, n (%)					
high school education level or less	75 (38.7)	57 (59.4)	17 (14.5)	149 (36.6)	$\chi^2(4)$= 61.96; p<.0001
Technical school	70 (36.1)	18 (18.8)	37 (31.6)	125 (30.7)	
University/Graduate	48 (24.1)	18 (18.8)	63 (53.8)	129 (31.7)	

Employment status, n (%)					
Student	5 (2.6)	2 (2.1)	17 (14.5)	24 (5.9)	$\chi^2(10)= 48.04$, $p<.0001$
Working full/part time	95 (49.0)	43 (44.8)	69 (59.0)	207 (50.9)	
Unemployed	7 (3.6)	3 (3.1)	3 (2.6)	13 (3.2)	
Medical leave	13 (6.7)	1 (1.0)	2 (1.7)	16 (3.9)	
Housewife	59 (30.4)	27 (28.1)	16 (13.7)	102 (25.1)	
Retired	15 (7.7)	18 (18.8)	9 (7.7)	42 (10.3)	

Abbreviations: FM = Fibromyalgia; RA = Arthritis Rheumatoid; and G = Attended at the Gynecologist.

Table 2. FSQ scores and clinical data.

	FM group	RA group	G group	Total
	M(SD)/n (%)	*M(SD)/n (%)*	*M(SD)/n (%)*	*M(SD)/n (%)*
Years with pain	7.46 (7.8)	8.96 (9.39)	-	7.96 (8.37)
Years since diagnosis	3.61 (4.42)	8.93 (9.52)	-	5.36 (6.99)
FSQ total score [0-31]	21.86 (5.34) *	14.69 (6.77)	9.38 (5.29)	16.58 (7.83)
SS score [0-12]	9.33 (2.19) *	6.75 (3.22)	5.43 (2.8)	7.6 (3.15)
Fatigue [0-3]	2.3 (0.76)	1.73 (1)	1.22 (0.93)	1.86 (0.99)
Cognitive Dysfunction [0-3]	2.08 (0.87)	1.36 (1.1)	1.11 (0.91)	1.63 (1.03)
Sleep unrefreshed [0-3]	2.51 (0.77)	1.92 (1.18)	1.56 (1.05)	2.1 (1.05)
Abdomen Pain [0-1)]	0.76 (0.43)	0.52 (0.5)	0.58 (0.5)	0.65 (0.48)
Depression [0-1]	0.79 (0.41)	0.53 (0.5)	0.24 (0.43)	0.57 (0.5)
Headache [0-1]	0.88 (0.32)	0.69 (0.47)	0.72 (0.45)	0.79 (0.41)
WPI [0-19]	12.53 (3.98) *	7.94 (4.49)	3.96 (3.3)	8.89 (5.38)
Antidepressants consumption				

Total	98 (50.5)	2 (2.1)	6 (5.1)	106 (26.0)
SSRIs	16 (16.3)	2 (100)	4 (66.7)	22 (20.8)
SNRIs	72 (73.5)	0	0	72 (67.9)
Tricyclic antidepressants	10 (10.2)	0	0	10 (9.4)
Atypical antidepressants	0 (0.0)	0	2 (33.3)	2 (1.9)

Note: *FM group presented a significantly higher score than the other groups (p <.0001).

Abbreviations: FSQ = Fibromyalgia Survey Questionnaire; FM = Fibromyalgia; RA = Rheumatoid Arthritis; G = Attended at the Gynecologist; SS = Symptom Severity Scale; WPI = Widespread Pain Index; SSRIs: Selective serotonin reuptake inhibitors, and SNRIs = Serotonin and norepinephrine reuptake inhibitors.

Table 3

Table 3. Descriptive statistics of instruments used.

	FM group					RA group					G group				
Instrument	Mean	SD	Min-Max	ICR		Mean	SD	Min-Max	ICR		Mean	SD	Min-Max	ICR	
				α	ω				α	ω				α	ω
FIQ-R	63.78	20.03	5.5-100	.93	.94	-	-	-	-	-	-	-	-	-	-
NPRS	23.63	6.53	0-38	.83	.83	18.28	8.61	0-37	.88	.88	10.09	1.47	0-31	.87	.88
PCS	30.34	13.93	0-52	.95	.96	-	-	-	-	-	-	-	-	-	-
PVAQ	48.89	13.28	15-79	.84	.88	-	-	-	-	-	-	-	-	-	-
PHQ-15	32.34	4.90	19-45	.77	.78	-	-	-	-	-	-	-	-	-	-
SF-12	35.21	7.09	21-56	.84	.86	-	-	-	-	-	-	-	-	-	-
BPI	6.43	2.24	0-10	.91	.91	-	-	-	-	-	-	-	-	-	-
PHQ-9	15.16	6.39	0-27	.84	.84	-	-	-	-	-	-	-	-	-	-

Abbreviations: FM = Fibromyalgia; RA = Arthritis Rheumatoid; and G = Attended at the Gynecologist; ICR = Internal Consistency Reliability; α = Cronbach's alpha; ω = McDonald's omega; FIQ-R = Fibromyalgia Impact Questionnaire, revised version; NPRS =

Numerical Pain Rating Scale; PCS = Pain Catastrophizing Scale; PVAQ = Pain Vigilance and Awareness Questionnaire; PHQ-15 = Patient Health Questionnaire-15; SF-12 = Short-Form Health Survey; BPI = Brief Pain Inventory, and PHQ-9 = Patient Health Questionnaire-9.

Table 4. *Structure matrix correlations of Hierarchical Discriminant Analysis.*

	FM versus Control (RA + G)		FM versus RA	
Variables	Step 1: Model including only the covariates	Step 2: Model adding FSQ total score	Step 1: Model including only the covariates	Step 2: Model adding FSQ total score
FSQ total Score	-	.90*	-	.72*
Age	-.20	.08	.43*	-.25
High-school or less vs Technical studies	-.33*	.14	-.40*	.23
High-school or less vs University/graduated studies	.27	-.11	.02	-.01
Married/live-in partner vs being single	.33*	-.13	.33*	-.19
Married/ live-in partner vs being divorced	-.01	.01	-.08	.05
Married/living vs being widowed	.54*	-.22	.57*	-.33*
Working vs being a student	.38*	-.16	-.05	.03

Working vs being unemployed	.23	-.10	-.14	.08
Working vs in medical leave	-.36*	.15	-.05	.03
Working vs being housewife	-.18	.07	-.19	.11
Working vs being retired	.36*	-.15	.44	-.26

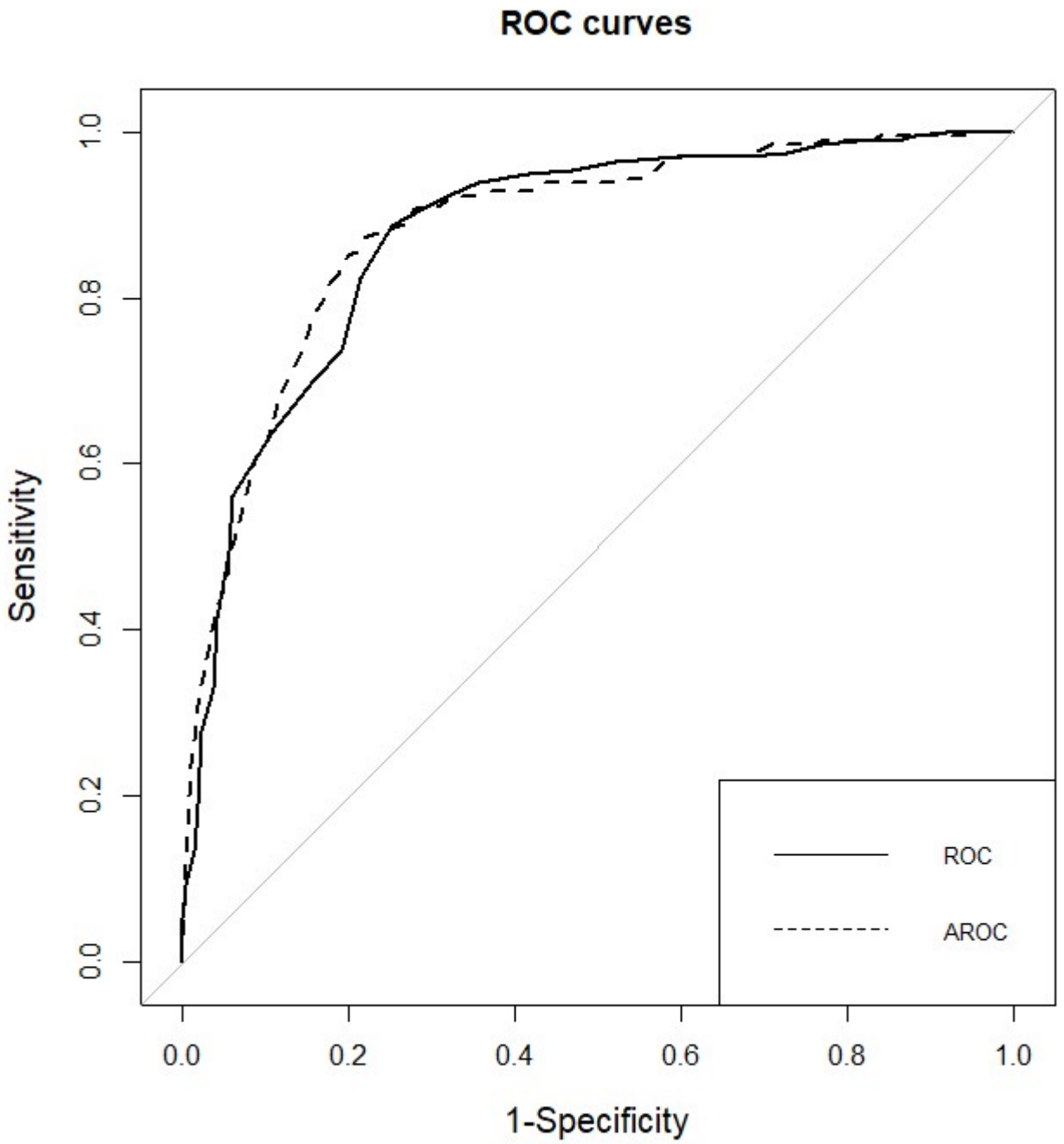
Note: * Discriminant loading higher than .30. *Abbreviations:* FM = Fibromyalgia; RA = Arthritis Rheumatoid; and G = Attended at the Gynecologist; FSQ = Fibromyalgia Survey Questionnaire.

Table 5. *Properties of PSD Cutoff Score selected*

Cutoff Score	Sensitivity	Specificity	PPV	NPV	PLR	NLR
12	.95	.53	.65	.93	2.03	.09
13[a]	.95	.58	.67	.93	2.27	.09
16	.91	.71	.74	.89	3.12	.13
17	.89	.75	.76	.88	3.50	.15

Note: [a] Cutoff reported by Wolfe et al., 2011 and Ferrari & Rusell, 2013. Abbreviations: PPV = Positive Predictive Value, NPV = Negative Predictive Value; PLR = Positive Likehood Ratio; and NLR = Negative Likehood Ratio.

Figure 1


ROC curves
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0
ROC
AROC