Psychometric characteristics of the Italian version of the revised Fibromyalgia Impact Questionnaire using classical test theory and Rasch analysis

F. Salaffi[1], F. Franchignoni[2], A. Giordano[3], A. Ciapetti[1], P. Sarzi-Puttini[4], M. Ottonello[5]

[1]Department of Rheumatology, Politechnic University of the Marche, Ancona, Italy; [2]Unit of Occupational Rehabilitation and Ergonomics, Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno (Novara), Italy; [3]Unit of Bioengineering, Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno (Novara), Italy; [4]Unità Operativa di Reumatologia, Ospedale L. Sacco, Milan, Italy; [5]Department of Physical and Rehabilitation Medicine, Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Nervi (Genoa), Italy.

Fausto Salaffi, MD, PhD
Franco Franchignoni, MD
Andrea Giordano, PhD
Alessandro Ciapetti, MD
Piercarlo Sarzi-Puttini, MD
Marcella Ottonello, PT, PhD

Please address correspondence to:
Fausto Salaffi, MD, PhD,
Department of Rheumatology,
Università Politecnica delle Marche,
Ospedale "C. Urbani",
Via dei Colli 52,
60035 Jesi (AN), Italy.
E-mail: fsalaff@tin.it

Received on January 10, 2013; accepted in revised form on April 8, 2013.

Clin Exp Rheumatol 2013; 31 (Suppl. 79): S41-S49.

Key words: fibromyalgia, psychometrics, Fibromyalgia Impact Questionnaire Revised version, Classical Test Theory, Rasch analysis

Competing interests: none declared.

ABSTRACT

Objectives. *The aim of this study was to perform a psychometric analysis of the Italian Fibromyalgia Impact Questionnaire Revised version (FIQR), using both classical test theory (CTT) and Rasch analysis (RA) in order to better analyse its construct validity and provide a rational basis for a possible improvement of its metric quality.*

Methods. *The study involved 503 patients with fibromyalgia (FM) (423 women and 80 men) with mean age of 51.3±10.1 years (range 19–74) and mean duration of symptoms of 11.1±8.7 years (range 1–30). All patients completed the Italian FIQR during their clinical visit. The translation and cultural adaptation process of the Italian FIQR followed the published guidelines and no local adjustments were made except for a slight adaptation of item 13 related to "energy".*

Results. *Factor Analysis revealed two salient dimensions: function (items 1–9) and symptoms (items 12–21). RA was thus performed on these two subscales. Rating scale diagnostics suggested collapsing the eleven rating categories of the scale into five. After combining these rating categories, RA showed that most items of each of the two subscales fitted the respective constructs to measure (MnSq 0.7–1.3). The reliability levels of the two subscales were higher than 0.80.*

Conclusion. *This study provides psychometric evidence of the reliability, internal validity and two-dimensional structure of the FIQR in a FM population. Our results support the use of two separate subscales for "function" and "symptoms", and provide a useful starting point for further refinement of the scale.*

Introduction

Fibromyalgia (FM) is a chronic multisymptom disease (1-3) affecting approximately 2–3% of the general population (more than 90% of patients are female), with pain as possibly its most important symptom (1, 4, 5). It consequently tends to have a profound impact on health-related quality of life (HRQL) (6) and is associated with high rates of healthcare resource use and an increased risk of inability to work (7), suicide and suicide attempts (8). Patients with FM report disability in daily living activities that is as severe as that reported by patients with rheumatoid arthritis (6, 9), and more severe than that reported by patients with osteoarthritis (10) or other painful conditions (11-14).

It has been shown that measuring HRQL is a key area of assessment when screening for disability and improving communication between patients and clinicians, and it is considered to be essential by regulatory agencies when contemplating the approval of medications for chronic pain status (15, 16).

The instruments used to measure physical function and health status are generally divided into generic and specific measures (17, 18): while the former provide a broad picture of health status over a range of conditions, the latter are more sensitive to the specific disorder and therefore more likely to reflect clinically important changes (18).

The Fibromyalgia Impact Questionnaire (FIQ) (19) and its revised version (FIQR) (20) are the most widely used FM-specific measures to assess the full spectrum of problems related to FM and its response to therapy. The FIQ was in the past recommended as a primary efficacy endpoint measure in FM clinical trials (21), and long considered to be the standard measure of multi-dimensional function/HRQL in FM patients. However, since questions were raised about its scoring (response options) and content validity (absence

of questions related to FM symptoms such as cognitive dysfunction, tenderness, balance problems, and environmental sensitivity), a revised version (FIQR) was developed to address these limitations without losing any of the essential properties of the original (22). The FIQR is currently the recommended means for assessing function and HRQL in FM patients (23). The psychometric properties of FIQR have been tested using only the classical test theory approach and focusing only on basic statistics, reliability item analysis, Cronbach's alpha and external construct validity (22, 24, 25). To our knowledge, no study has analysed the FIQR using Rasch analysis, a statistical method for testing if the properties of a questionnaire comply with a wide range of psychometric requirements that cannot be analysed using classical test theory.

The aim of this study was to perform a comprehensive psychometric analysis of the Italian version of the FIQR in a large sample of FM patients, using both classical test theory (CTT) and Rasch analysis, in order to examine its main metric properties (*i.e.* rating scale functioning, internal construct validity, reliability indices, dimensionality) and provide insights for a possible improvement of the questionnaire.

Patients and methods

Study population

The study was carried out in 503 FM patients (423 women and 80 men), mean age 51.3±10.1 years (range 19–74) and mean duration of symptoms 11.1±8.7 years (range 1–30). All patients were recruited from the Department of Rheumatology of the Politechnic University of the Marche, Ancona, Italy, and were considered eligible if they were at least 18 years old and met the criteria for fibromyalgia defined by the American College of Rheumatology (ACR) (26). The main exclusion criteria were: an unstable medical or psychiatric illness or current primary psychiatric diagnosis including severe depression, pain due to traumatic injury or structural/regional rheumatic disease, rheumatoid arthritis, inflammatory arthritis, and autoimmune disease.

Patients were examined for FM symptoms and clinical findings in order to confirm the diagnosis, and then they were asked to complete the Italian FIQR.

The disease-related symptoms were also evaluated by the Fibromyalgia Assessment Status (FAS) (27). Briefly, the FAS is a short, self-administered

VERSIONE ITALIANA DEL REVISED FIBROMYALGIA IMPACT QUESTIONNAIRE (ITALIAN-FIQR)

Nome: _____ Età: _____

Cognome: _____ Durata dei sintomi (anni): _____

DOMINIO 1 – FUNZIONE FISICA

Per ognuna delle seguenti NOVE domande, segni la casella che meglio indica il grado di difficoltà da Lei avvertito, ***nel corso dell'ultima settimana,*** nel compiere ciascuna delle attività elencate, a causa della fibromialgia: ***(scegliere un solo numero).***

1. SPAZZOLARE O PETTINARE I CAPELLI

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

2. CAMMINARE ININTERROTTAMENTE PER 20 MINUTI

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

3. PREPARARE I PASTI

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

4. PASSARE L'ASPIRAPOLVERE E LAVARE I PAVIMENTI

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

5. SOLLEVARE E PORTARE LE BORSE DELLA SPESA

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

6. SALIRE UN PIANO DI SCALE

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

7. CAMBIARE LE LENZUOLA DEL LETTO

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

8. STARE SEDUTO SU UNA SEDIA PER ALMENO 45 MINUTI

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

9. ANDARE A FARE LA SPESA

Nessuna difficoltà	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Estrema difficoltà

DOMINIO 1 – SUBTOTALE:_____

DOMINIO 2 – STATO DI SALUTE GENERALE

Per ognuna delle seguenti DUE domande, segni la casella che meglio indica il grado di difficoltà da Lei avvertito, ***nel corso dell'ultima settimana,*** nel compiere ciascuna delle attività elencate, a causa della fibromialgia: ***(scegliere un solo numero).***

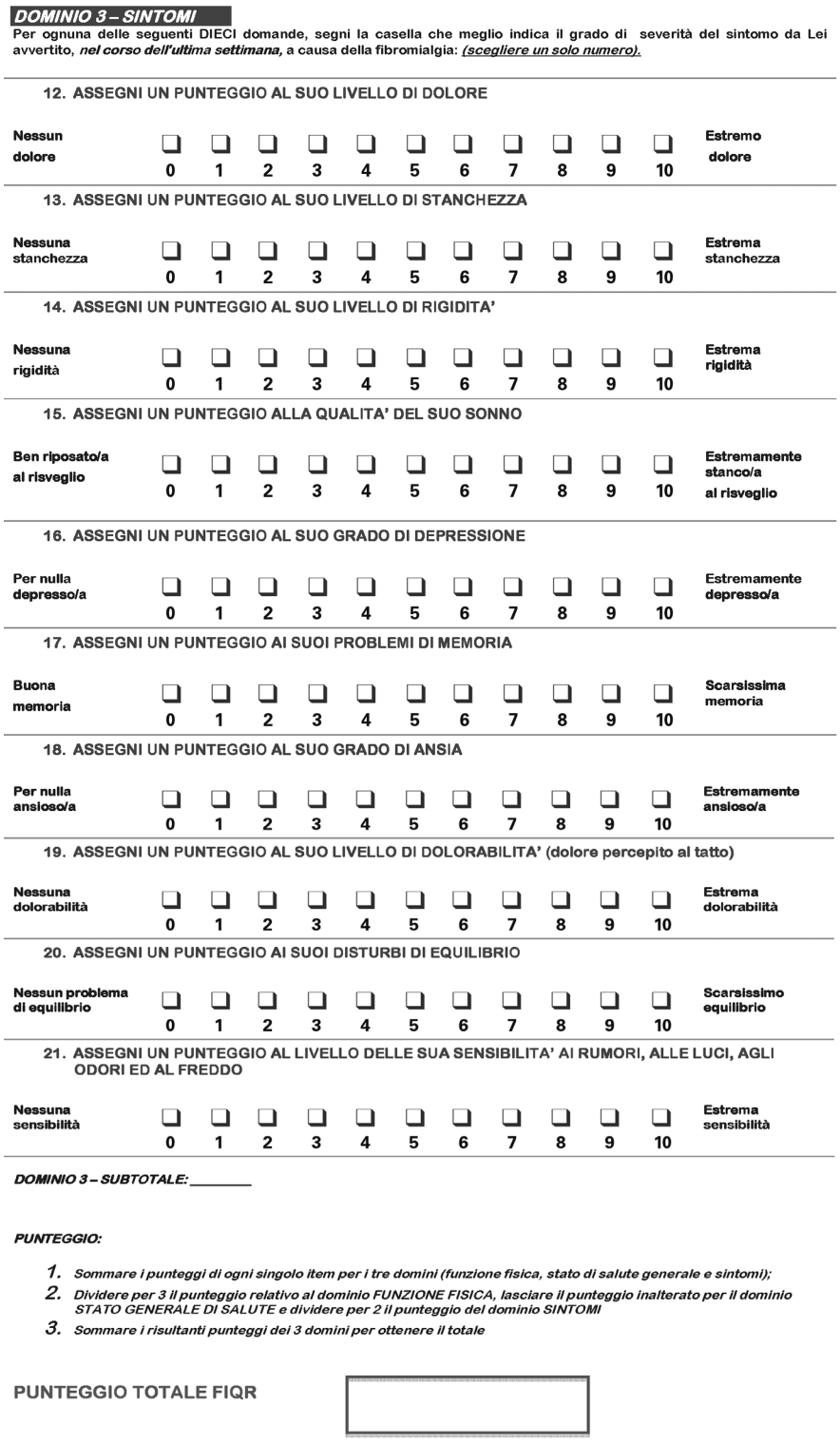
10. LA FIBROMIALGIA MI HA IMPEDITO DI PORTARE A TERMINE I LAVORI/COMPITI DELLA SETTIMANA

Mai	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Sempre

11. SONO STATO/A COMPLETAMENTE SOPRAFFATTO/A DAI SINTOMI DELLA FIBROMIALGIA

Mai	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6	❑ 7	❑ 8	❑ 9	❑ 10	Sempre

DOMINIO 2 – SUBTOTALE:_____



DOMINIO 3 – SINTOMI

Per ognuna delle seguenti DIECI domande, segni la casella che meglio indica il grado di severità del sintomo da Lei avvertito, *nel corso dell'ultima settimana,* a causa della fibromialgia: *(scegliere un solo numero).*

12. ASSEGNI UN PUNTEGGIO AL SUO LIVELLO DI DOLORE

Nessun dolore ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estremo dolore

13. ASSEGNI UN PUNTEGGIO AL SUO LIVELLO DI STANCHEZZA

Nessuna stanchezza ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estrema stanchezza

14. ASSEGNI UN PUNTEGGIO AL SUO LIVELLO DI RIGIDITA'

Nessuna rigidità ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estrema rigidità

15. ASSEGNI UN PUNTEGGIO ALLA QUALITA' DEL SUO SONNO

Ben riposato/a al risveglio ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estremamente stanco/a al risveglio

16. ASSEGNI UN PUNTEGGIO AL SUO GRADO DI DEPRESSIONE

Per nulla depresso/a ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estremamente depresso/a

17. ASSEGNI UN PUNTEGGIO AI SUOI PROBLEMI DI MEMORIA

Buona memoria ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Scarsissima memoria

18. ASSEGNI UN PUNTEGGIO AL SUO GRADO DI ANSIA

Per nulla ansioso/a ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estremamente ansioso/a

19. ASSEGNI UN PUNTEGGIO AL SUO LIVELLO DI DOLORABILITA' (dolore percepito al tatto)

Nessuna dolorabilità ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estrema dolorabilità

20. ASSEGNI UN PUNTEGGIO AI SUOI DISTURBI DI EQUILIBRIO

Nessun problema di equilibrio ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Scarsissimo equilibrio

21. ASSEGNI UN PUNTEGGIO AL LIVELLO DELLE SUA SENSIBILITA' AI RUMORI, ALLE LUCI, AGLI ODORI ED AL FREDDO

Nessuna sensibilità ❑ 0 ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8 ❑ 9 ❑ 10 Estrema sensibilità

DOMINIO 3 – SUBTOTALE:_____

PUNTEGGIO:

1. *Sommare i punteggi di ogni singolo item per i tre domini (funzione fisica, stato di salute generale e sintomi);*
2. *Dividere per 3 il punteggio relativo al dominio FUNZIONE FISICA, lasciare il punteggio inalterato per il dominio STATO GENERALE DI SALUTE e dividere per 2 il punteggio del dominio SINTOMI*
3. *Sommare i risultanti punteggi dei 3 domini per ottenere il totale*

PUNTEGGIO TOTALE FIQR

Fig. 1. Italian version of the Fibromyalgia Impact Questionnaire Revised version (FIQR).

index combining questions related to non-articular pain (range 0–10), fatigue (range 0–10) and quality of sleep (range 0–10), that provides a single composite measure of disease activity. The total score is calculated by summing the three subscores and dividing the result by three (range 0–10). The FAS has demonstrated validity, reliability and responsiveness (27, 28).

Our local Human Ethics Committees approved the research protocol and all patients gave their written informed consent.

Translation and cultural adaptation of the FIQR

The Fibromyalgia Impact Questionnaire Revised version (FIQR) is an updated version of the FIQ (19) that was developed by Bennett *et al*. (20) to address the limitations of the original FIQ (22). The new version has 21 items (all based on an 11-point numeric rating scale of 0–10, with higher scores indicating greater disease impact), and covers the three domains of function, overall impact, and symptoms. FIQR is an improvement on the former scale mainly due to the addition of new questions related to memory, tenderness, balance, and environmental sensitivity. As in the FIQ, all questions refer to the previous seven days, and the total maximum score is 100, with higher scores indicating greater disease impact (20). The FIQR score is the sum of the three domain scores: the summed score for the 9-item function domain (range 0–90) is divided by three; the summed score for the 2-item overall impact domain (range 0–20) remains as it is; and the summed score for the 10-item symptom domain (range 0–100) is divided by two.

The translation and cultural adaptation process followed established guidelines (29). The original English version of the FIQR was translated into Italian by two independent professional translators (Fig. 1). One translator was aware of the questionnaire, the other was not. The two forward translations were reviewed and compared with one another and with the original English version by the translators and one of the authors. Subsequently, a pilot test was run including a cognitive debriefing (*i.e*. an interview by a psychologist about the clarity, intelligibility, appropriateness, and cultural relevance of the target language version), and the final adaptation was prepared by experts. No local adjustments were made, except for a slight adaptation of item 13 related to "energy", in which the descriptors are "0=no fatigue" and "10=extreme fatigue". This was because the word "fatigue" is better received by Italian patients, and it also represents one of the major domains classified as important by patients (30).

Statistical analysis

We combined the approaches of CTT and Rasch analysis in order to investigate the psychometric properties of FIQR. In accordance with CCT, the internal consistency of the FIQR was assessed using Cronbach's coefficient alpha (values > 0.85–0.90 are desirable for individual judgments), and item-to-total correlation (the usual rule of thumb is that an item should correlate with the total score with r>0.20) (31). Considering the unknown factorial structure of the responses to the FIQR, an estimate of the number of relevant factors was obtained by means of parallel analysis (PA) (32), and then an exploratory factor analysis (EFA) for ordinal data (33) with oblique (Promax) rotation on a randomly split half of the dataset (n=242) was used to study the contribution of each item to the factors identified by the PA. An item was considered as contributing to a factor when factor loading was 0.4 or more: this threshold was chosen taking into account sample size, number of items and the exploratory nature of the study (34). As mono-dimensionality was not confirmed, the number of underlying factors and their relation to each item was used to separate the scale into subscales, the clinical acceptability of which was evaluated by experts. Each subscale underwent further analyses after verifying a good fit between the data and the model using confirmatory factor analysis (CFA) on the second half of the dataset (n=243). The following indices were used to evaluate the fit of the data to the model: the Tucker-Lewis fit index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR). For an acceptable fit, TLI and CFI should be ≥0.95, RMSEA ≤0.08 and SRMR ≤0.10 (35).

Once the factorial structure of the scale was determined, the matrix of the individual raw scores of each subject (n=503) on each of the two selected subscales underwent Rasch analysis using WINSTEPS software v. 3.68.2 (36). The technical aspects of the analysis have been described elsewhere (37, 38). As a first step, we used a rating scale model to investigate whether the rating scale of the FIQR was being used in the expected manner. The response categories were evaluated and collapsed using standardised procedures (39). After rating scale modifications, a new series of Rasch analyses of the two subscales were performed as suggested by the preliminary dimensionality analysis. The validity of each scale was assessed by evaluating the fit of the individual items to the latent trait in accordance with the Rasch model and examining whether the pattern of item difficulties was consistent with the model expectations. Information-weighted (infit) and outlier-sensitive (outfit) mean-square statistics (MnSq) for each item were calculated, considering a MnSq value between 0.8 and 1.2 as indicating an acceptable fit (37). Reliability was evaluated in terms of 'separation', which was defined as the ratio between the true spread of the measures and their measurement error. A separation of 2.0 is considered good, and a related index is the reliability of these separation indices, which provides the degree of confidence that can be placed in the consistency of the estimates (range 0–1; coefficients of >0.80 are considered good, and >0.90 excellent) (37). A principal component analysis (PCA) of the standardised residuals was used to investigate the local independence of the items (36) and the presence of sub-dimensions as an independent confirmation of the mono-dimensionality of the scale. The following criteria were used to confirm mono-dimensionality: a) a cut-off of 50% of the variance explained by the trait that the scale is intended to measure (the 'Rasch factor'); and b) the eigenvalue of the first residual factor smaller than three (36).

In addition, convergent validity of the two FIQR subscales ("function"; "symptoms") was assessed by calculat-

Table I. Scores obtained on the FIQR and FAS (mean and median values, standard deviations and 95% confidence intervals for each item, sub-dimensions and total score of the FIQR, and for total score of the FAS) by the study patients (n=503).

Item n	Item description	Mean	Median	Standard deviation	95% confidence interval	Score range
1	Brush or comb hair	2.48	1.00	2.77	2.24–2.72	0–10
2	Walk continuously for 20 minutes	4.54	5.00	3.00	4.27–4.80	0–10
3	Prepare a homemade meal	3.28	3.00	2.76	3.03–3.52	0–10
4	Vacuum, scrub or sweep floors	4.67	5.00	2.98	4.41–4.93	0–10
5	Lift and carry a bag full of groceries	5.32	5.00	2.83	5.07–5.57	0–10
6	Climb one flight of stairs	4.33	5.00	3.09	4.06–4.61	0–10
7	Change bed sheets	4.04	4.00	2.92	3.78–4.30	0–10
8	Sit in a chair for 45 minutes	4.49	5.00	3.08	4.21–4.76	0–10
9	Go shopping for groceries	3.86	4.00	2.88	3.61–4.12	0–10
	FIQR function sub-total	*12.32*	*12.00*	*6.90*	*11.72–12.93*	*0–30*
10	Cannot achieve goals	4.72	5.00	2.76	4.48–4.96	0–10
11	Feel overwhelmed	3.26	3.00	3.15	2.98–3.54	0–10
	FIQR overall impact sub-total	*7.98*	*7.00*	*5.24*	*5.52–8.44*	*0–20*
12	Pain rating	5.94	6.00	2.10	5.76–6.13	0–10
13	Fatigue rating	6.42	7.00	2.36	6.21–6.63	0–10
14	Stiffness rating	5.76	6.00	2.21	5.57–5.96	0–10
15	Sleep quality	5.82	6.00	2.63	5.59–6.05	0–10
16	Depression level	4.49	5.00	2.46	4.27–4.70	0–10
17	Memory problems	5.25	5.00	2.58	5.02–5.48	0–10
18	Anxiety level	5.32	6.00	2.58	5.09–5.54	0–10
19	Tenderness level	4.90	5.00	2.50	4.68–5.12	0–10
20	Balance problems	4.22	4.00	2.39	4.01–4.43	0–10
21	Environmental sensitivity	5.72	6.00	2.56	5.49–5.94	0–10
	FIQR symptoms sub-total	*27.17*	*27.00*	*8.29*	*26.45–27.90*	*0–50*
	FIQR TOTAL SCORE	*47.32*	*47.00*	*18.09*	*45.73–48.90*	*0–100*
	FAS TOTAL SCORE	*5.74*	*5.71*	*1.71*	*5.59–5.89*	*0–10*

FIQR: Fibromyalgia Impact Questionnaire Revised version; FAS: Fibromyalgia Assessment Status.

ing the Pearson correlation coefficient (*r*) between their raw scores and that of the FAS. STATA 10.1 (StataCorp, LP College Station, TX, USA) was used to perform the PA. Lisrel 8.80 (Scientific Software International, Inc. Lincolnwood, IL, USA) was used for the EFA and CFA.

Results

Descriptive statistics

Table I shows the score details for each item, subscales and total score of the FIQR, as well as for the FAS. Frequency distribution of FIQR scores was simmetrical, with a positive excess kurtosis (mean 47.32; median 47; skewness 0.20; kurtosis 2.54). The spydergram in Figure 2 shows the distribution of the FIQR scores in the patient sample as a whole: it offers a simplified means to visualise complex results across all domains of FIQR in a single figure. The five highest scoring items (greater disease impact) were related to symptoms: fatigue/energy (FIQR13), pain (FIQR12), stiffness (FIQR14), sleep quality (FIQR15), and environmental sensitivity (FIQR21). The lowest scoring items included functional activities, such as brushing/combing hair, preparing a home-made meal, shopping for groceries, and changing bed sheets.

Classical test theory analysis

The Cronbach alpha of the FIQR was 0.94. All items had an item-to-total correlation between 0.41 (item 21) and 0.78 (item 9). After randomly splitting the dataset into two halves, PA applied to the first half revealed two factors with empirical eigenvalues that exceeded those averaged from 200 random datasets with the same number of variables and observations, that respectively explained 57.9% and 8.6% of the variance (66.5% cumulatively). In the subsequent two factor EFA, results of which are presented in Table II, items 10 and 11, supposed to represent "overall impact", failed to achieve salient loadings.

On the basis of these findings (consistent with expert opinion), two separate subscales were identified: the first including FIQR items 1–9 ("function"), the second composed of items 12–21 ("symptoms"). Items 10 and 11 were therefore not analysed further.

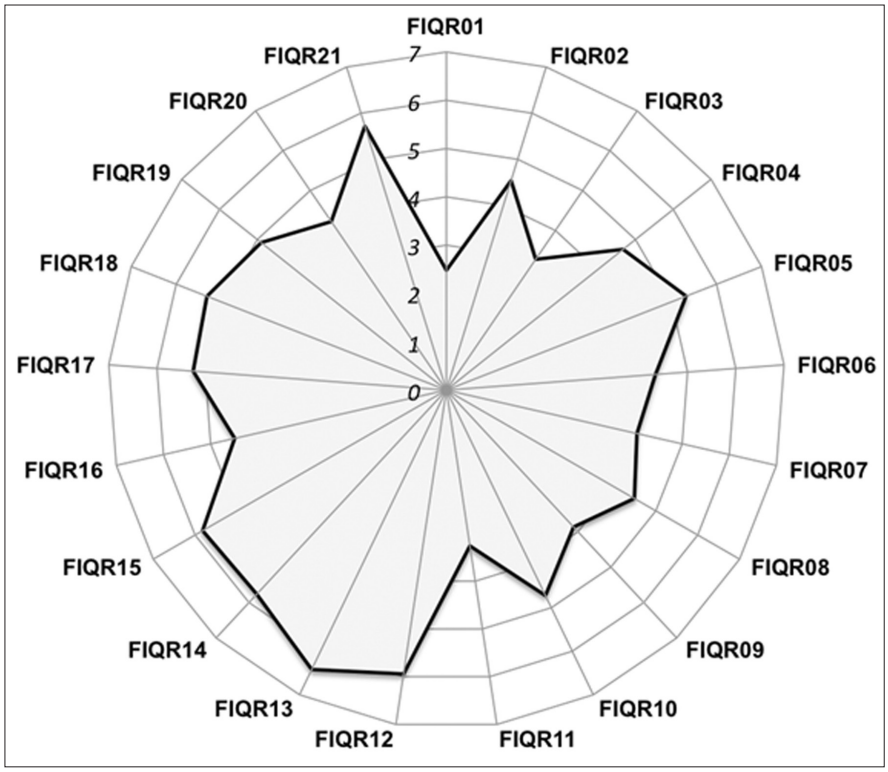


Fig. 2. Spydergram of the FIQR domains of the overall cohort of fibromyalgia patients (n=503). The domain scores are plotted from 0 (worst, at the centre) to 10 (best, at the outside). For item numbers, see Tables I and II.

A CFA of the 2-factor model on the second half of the dataset showed a borderline fit: TLI, CFI, RMSEA, and SRMR were respectively 0.94, 0.95, 0.11, and 0.068.

As expected, a good correlation was found between the FAS total score and both FIQR subscales (r=0.64 for "function"; r=0.84 for "symptoms"). The correlation between FIQR and FAS was good (r=0.78). The correlations between FIQR and FAS subdomais (pain, fatigue and sleep) were moderate and good (r=0.46, r=0.65 and r=0.61, respectively).

Rasch analysis

Rating scale diagnostics showed that the rating categories did not comply with the pre-set criteria for category functioning (average measures, thresholds, etc.), suggesting that raters had difficulty in discerning the eleven response levels proposed by the original FIQR. Accordingly, the response categories were revised: the pre-set criteria were best met by collapsing categories 1–3, 4–6, and 7–9 (Fig. 3). As shown in Table II, after combining these rating scale categories, all the items of each of the subscales ("function" and "symptoms") fitted their respective constructs (MnSq 0.8–1.2) except: a) the underfitting item 1 "Brush or comb hair" in the function subscale (Infit Mnsq=1.44; Outfit Mnsq=1.29); b) the borderline items 17 "Memory problems" (Infit Mnsq=1.21; Outfit Mnsq=1.23) and 21 "Sensitivity to loud noises, bright lights, odours and cold" (Infit Mnsq=1.21; Outfit Mnsq=1.13) in the symptom subscale; and c) the slightly overfitting item 3 "Prepare a homemade meal" (Infit Mnsq=0.78; Outfit Mnsq=0.76) and item 9 "Go shopping for groceries" (Infit Mnsq=0.75; Outfit Mnsq=0.70) in the function subscale, and item 14 "Level of stiffness" in the symptom subscale (Infit Mnsq=0.76; Outfit Mnsq=0.75). Table III summarises the results for subject ability, item difficulty, reliability indices, and the two PCAs of the standardised residuals. The sample-item matching was very good for "Symptoms" (average subject ability=0.03) and acceptable for "Function" (average sub-

Table II. Exploratory factor analysis of a two-factor solution as suggested by the parallel analysis (loadings of >0.40 are in bold) (sample size: n=242) and Rasch analysis with item calibration and fit information (under-fitting values in bold and over-fitting values in italics; sample size: n=503). Items with higher (positive) Measure in Rasch analysis are those showing lower level of difficulty.

Item	Factor analysis		Rasch analysis		
	Factor 1	Factor 2	Measure (SE)	Infit MnSq	Outfit MnSq
1 Brush or comb hair	**0.638**	-0.029	1.20 (0.07)	**1.44**	**1.29**
2 Walk continuously for 20 minutes	**0.680**	0.168	-0.31 (0.06)	0.86	0.83
3 Prepare a homemade meal	**0.894**	-0.095	0.61 (0.06)	*0.78*	*0.76*
4 Vacuum, scrub or sweep floors	**0.722**	0.011	-0.41 (0.06)	1.16	1.11
5 Lift and carry a bag full of groceries	**0.665**	0.134	-0.88 (0.06)	0.81	0.79
6 Climb one flight of stairs	**0.460**	0.387	-0.16 (0.06)	1.11	1.07
7 Change bed sheets	**0.737**	0.031	0.03 (0.06)	0.96	0.90
8 Sit in a chair for 45 minutes	**0.823**	-0.125	-0.27 (0.06)	1.15	1.19
9 Go shopping for groceries	**0.701**	0.174	0.19 (0.06)	*0.75*	*0.70*
10 Cannot achieve goals	0.395	0.378	–	–	–
11 Feel overwhelmed	0.288	0.343	–	–	–
12 Pain rating	0.140	**0.583**	-0.48 (0.07)	0.87	0.86
13 Fatigue rating	-0.026	**0.736**	-0.80 (0.07)	1.02	1.00
14 Stiffness rating	0.146	**0.629**	-0.24 (0.07)	*0.76*	*0.75*
15 Sleep quality	-0.088	**0.739**	-0.30 (0.07)	1.03	1.05
16 Depression level	0.295	**0.417**	0.66 (0.06)	1.03	1.07
17 Memory problems	-0.029	**0.530**	0.06 (0.07)	**1.21**	**1.23**
18 Anxiety level	0.011	**0.668**	0.05 (0.07)	1.07	1.03
19 Tenderness level	0.243	**0.409**	0.39 (0.06)	1.02	1.09
20 Balance problems	0.213	**0.541**	0.83 (0.06)	0.80	0.81
21 Environmental sensitivity	-0.111	**0.598**	-0.19 (0.07)	**1.21**	**1.13**

ject ability=-0.90). Separation indexes and respective reliability were satisfactory for both subscales (person separation reliabilities ≥0.83; item separation reliabilities ≥0.98). Subject ability span was good (>11 logits in each subscale), whereas item difficulty span ranged from 1.65 logits (symptoms) to 2.08 logits (function). The two PCAs on the residuals did not sustain the presence of additional dimensions, thus confirming the unidimensionality of each subscale, while the absence of residual correlations >0.30 between item pairs verified the local independence of their items.

Discussion

Patient-reported outcome measures such as the FIQR are increasingly being used as a key part in the evaluation of clinical health care interventions, but their ability to improve decision making greatly relies on their psychometric strength (18). Bennet *et al*. proposed the FIQR in 2009 as an extensive modification of the previous questionnaire in terms of content, number of items and response options (20). It was designed to overcome some of the drawbacks with the FIQ (22), but its validation (and that of the subsequent translations and cultural adaptations) used only a classical test theory approach.

New and more detailed psychometric approaches, including a mix of classical test theory and item response theory methods, have recently been recommended to provide reliability and validity estimates of the instrument (40). This study is the first psychometric analysis of FIQR using both CTT and Rasch analysis to examine its properties in depth and provide a rationale for improving its metric quality. As such we mainly concentrated on analysing its dimensions and rating scale diagnostics, and identifying the appropriateness of its items for measuring the intended construct.

The five most endorsed (*i.e*. highest scoring) items are all related to symptoms: fatigue/energy, pain, stiffness, sleep quality, and environmental sensitivity. These findings are very similar to those reported by Bennett *et al*. (20) and Ediz *et al*. (25), thus supporting the importance of these domains for a patient-centred evaluation (15, 30, 41).

Cronbach's alpha and the item-to-total statistics showed acceptable values according to CTT rules. This finding is similar to that of Bennett *et al*. (Cronbach's alpha 0.95) (20), and more favourable than that of Ediz *et al*. (25) and Srifi *et al*. (24), who respectively reported alpha values of 0.89–0.91 and 0.91–0.92. However, in terms of monodimensionality, the factor analysis (parallel analysis and EFA) hinted at two factors: the first factor consisting of the first nine items (related to the "function" domain), the second consisting of the last 10 items (no. 12–21) related to the "symptoms" domain. These are two of the three domains defined by the original authors in creating the scale. The third domain – "overall impact" (consisting of questions no. 10 and 11) – did not emerge as an independent factor, which is in line with the observation that, under the best conditions, a minimum of three items is critical for producing a dimension in factor analysis (42). These three domains are often used separately to validate the scale (20, 24, 25) and, in terms of total score, the "function" domain accounts for 30%, "symptoms" for 50%, and "overall impact" for 20% (20). This weighting system seems to be based on theoretical considerations (expert evaluation resulting in different weighting of the three domains), rather than a statistical approach, and we think it warrants further research.

Rasch analysis of the 11 rating scale categories showed that the FIQR has many disordered thresholds. The category thresholds are the ability levels at which the response to either of the two adjacent categories is equally likely. Disordered thresholds occur when respondents have difficulty consistently discriminating between response options (*e.g*. because there are too many response options or the labelling of the options is confusing). Our findings indicate that our subjects were able to discern only five categories (rather than the original eleven levels), and hence it might be appropriate to simplify the item format.

This is not the first time that an 11-level numeric rating scale (0–10) has failed

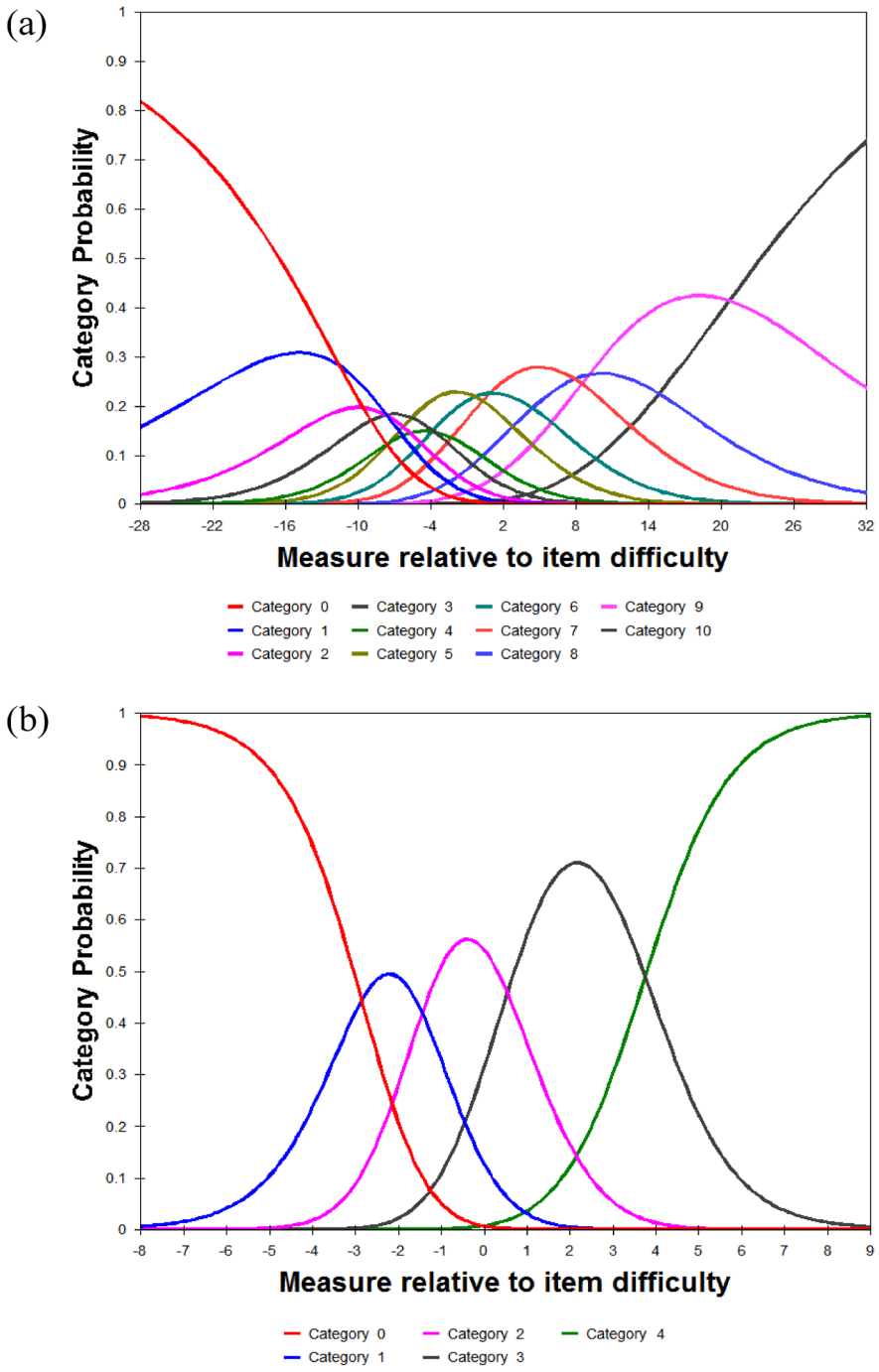


Fig. 3. Category probability curves: **a**) Curves of the original 11 categories (0–10); and **b**) of the 5 revised categories obtained by combining the original categories as follows: 0=0; 1–3=1; 4–6=2; 7–9=3; 10=4. The y axis represents the probability (0–1) of responding to one of the rating categories and the x axis the different performance values (patient ability minus item difficulty) in logits. The ideal plot should look like an ordered even succession of hills, with an 'emerging' crest where each category is modal over a certain range. The "0" curve declines as the subject's ability increases; the crossing point (where 0 and 1 are equally probable) is the first "threshold". The same applies to the other curves. The graph in (**a**) shows that the probability of using some categories is never higher than that of other adjacent ratings, whereas the graph in (**b**) shows that the probability of selecting each of the five revised rating categories is a clear function of the level of ability shown by the subject in the x axis.

to demonstrate appropriate functioning. For example, an analysis of the mobility section of the Prosthesis Evaluation Questionnaire (a self-administered questionnaire designed to evaluate subjects who have undergone amputation) showed that patients were unable to distinguish their abilities as finely as suggested by the 0–10 numeric rating scale, whereas five well-selected response categories were able to improve the measurement qualities of the scale without decreasing its reliability indices (43). Similarly, when Khadka *et al.* (44) used Rasch analysis to assess the performance of the rating scales of seventeen outcome measures in ophthalmology, they found that a simple and uniform question format (*e.g.* four or five well-labelled categories) was more likely to function (from a psychometric point of view) than those with a complicated question format and a large number of response categories.

As for the fit of the individual items to the latent trait (analysis of internal construct validity), the item "Brush or comb hair" (that shows the lower level of difficulty in the function subscale) in the function subscale was underfitting, which indicates either that it does not tap the same underlying construct as the other items, or that it is poorly written or, more probably, too sensitive to (personal and environmental) confounding factors. This item was added to the FIQR without any formal validation at item level, following a suggestion from the focus group. If our findings are confirmed, it might be opportune to consider replacing it with one that is more homogeneous with the scale's conceptual framework, and capable of optimising content validity, coverage and the technical quality of the measure in line with modern psychometric approaches. Moreover, the items "Memory problems" and "Sensitivity to loud noises, bright lights, odours and cold" (in the "symptoms" subscale) had a borderline fit. They warrant further investigation but, for the present, we suggest retaining both as the misfit is minor and they are clinically important. Similarly, the items "Prepare a homemade meal" and "Go shopping for groceries" in the "function" subscale and "Level of stiffness" in the "symptoms" subscale present a too predictable pattern at Rasch analysis (overfit) and could perhaps be deleted (as redundant) or substituted. Alternatively, they could be retained if considered clinically relevant because overfitting simply tends to overestimate differences in raw scores,

Table III. Subjects'ability levels and reliability indices for the two FIQR subscales (n=503).

	Subscale 1 "Function" (items 1–9)		Subscale 2 "Symptoms" (items 12–21)	
	Original coding	Recoded	Original coding	Recoded
Average ability levels (range)	-0.41 (-3.79–2.92)	-0.90 (-6.08–5.62)	0.03 (-4.16–2.73)	0.03 (-6.51–5.30)
SD	0.91	1.80	0.60	1.36
Person separation index	2.26	2.70	2.39	2.23
(person separation reliability)	(0.84)	(0.88)	(0.85)	(0.83)
Cronbach's alpha	0.92	0.91	0.87	0.85
Range of item difficulty estimates	-0.38 (item 5) to 0.53 (item 1)	-0.88 (item 5) to 1.20 (item 1)	-0.34 (item 13) to 0.36 (item 20)	-0.80 (item 13) to 0.85 (item 20)
SD	0.25	0.57	0.20	0.49
Item separation index	9.38	8.55	7.70	6.99
(item separation reliability)	(0.99)	(0.99)	(0.98)	(0.98)
Variance explained by the Rasch factor	62.9%	58.4%	49.0%	44.8%
Eigenvalue of the first residual factor	2.2	2.2	1.8	1.7
Variance explained by the first residual factor	9.0%	10.2%	9.3%	9.4%

without affecting dimensionality or severely degrading the metric properties of the scale.

The item difficulty range of both subscales (Table III) covers an acceptable distance of each targeted construct (between 1.6 and 2 logits). The targeting of item difficulty to patient ability (*i.e.* the extent to which the items are appropriately difficult for the sample) was very good in the case of 'symptoms' and acceptable in the case of 'function'. However, a study of the stability of item hierarchy (differential item functioning) across sub-samples defined on the basis of potentially relevant clinical criteria (*e.g.* age, gender or symptom patterns) is warranted.

A number of potential limitations of this study should be mentioned. No additional dimension was suggested by RA, but the CFA of the 2-factor model on the second half of the dataset showed only a borderline fit that could be due to the limited sample size. We cannot exclude that some specific (linguistic, cultural or technical) characteristics of the Italian version of the FIQR may have influenced some results, although this version was obtained by means of a thorough forward/backward translation process. A further caveat is that this analysis was based on a population of adults from a relatively limited geographical area of Italy. Finally, the present findings need to be confirmed in different countries, contexts and patient samples.

Conclusion

To our knowledge, this is the first study investigating the dimensionality and validity of the FIQR using stringent criteria that included item response theory methods.

Our main findings are:

a) two of the three domains were confirmed, whereas the third probably needs some more items in order to become an interpretable domain from a psychometric point of view;

b) the selection of the items in the two analysed domains seems to be appropriate (also in terms of coverage and technical quality), although there is room for some refinement;

c) the 0–10 rating scale should be simplified. A five-level response structure seems to be enough, but its actual performance needs to be analysed. This preliminary study supports the usefulness of this questionnaire for measuring the multidimensional spectrum of FM-related functional (45-47) and HRQL problems, and provides a basis for further research aimed at improving its measurement qualities.

References

1. WOLFE F, ROSS K, ANDERSON J, RUSSELL IJ, HEBERT L: The prevalence and characteristics of fibromyalgia in the general population. *Arthritis Rheum* 1995; 38: 19-28.
2. MEASE P: Fibromyalgia syndrome: review of clinical presentation, pathogenesis, outcome measures, and treatment. *J Rheumatol* 2005; 32 (Suppl. 75): 6-21.
3. SALAFFI F, SARZI-PUTTINI P: Old and new criteria for the classification and diagnosis of fibromyalgia: comparison and evaluation. *Clin Exp Rheumatol* 2012; 30 (Suppl. 74): S3-9.
4. SALAFFI F, DE ANGELIS R, GRASSI W; MARCHE PAIN PREVALENCE, INVESTIGATION GROUP (MAPPING) STUDY: Prevalence of musculoskeletal conditions in an Italian population sample: results of a regional community-based study. I. The MAPPING study. *Clin Exp Rheumatol* 2005; 23: 819-28.
5. MAS AJ, CARMONA L. M. VALVERDE M, RIBAS B, AND THE EPISER STUDY GROUP: Prevalence and impact of fibromyalgia on function and quality of life in individuals from the general population: results from a nationwide study in Spain. *Clin Exp Rheumatol* 2008; 26: 519-26.
6. SALAFFI F, SARZI-PUTTINI P, GIROLIMETTI R, ATZENI F, GASPARINI S, GRASSI W: Health-related quality of life in fibromyalgia patients: a comparison with rheumatoid arthritis patients and the general population using the SF-36 health survey. *Clin Exp Rheumatol* 2009; 27 (Suppl. 56): S67-74.
7. KIVIMÄKI M, LEINO-ARJAS P, KAILA-KANGAS L *et al*.: Increased absence due to sickness among employees with fibromyalgia. *Ann Rheum Dis* 2007; 66: 65-9.
8. CALANDRE EP, VILCHEZ JS, MOLINA-BAREA R *et al*.: Suicide attempts and risk of suicide in patients with fibromyalgia: a survey in Spanish patients. *Rheumatology* (Oxford) 2011; 50: 1889-93.
9. OFLUOGLU D, BERKER N, GÜVEN Z, CANBULAT N, YILMAZ IT, KAYHAN O: Quality of life in patients with fibromyalgia syndrome and rheumatoid arthritis. *Clin Rheumatol* 2005; 24: 490-2.
10. HAWLEY DJ, WOLFE F: Pain, disability, and

pain/disability relationships in seven rheumatic disorders: a study of 1,522 patients. *J Rheumatol* 1991; 18: 1552-7.
11. SALAFFI F, CAROTTI M, GASPARINI S, INTORCIA M, GRASSI W: The health-related quality of life in rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis: a comparison with a selected sample of healthy people. *Health Qual Life Outcomes* 2009; 18: 7-25.
12. YILMAZ F, SAHIN F, ERGOZ E *et al*.: Quality of life assessments with SF 36 in different musculoskeletal diseases. *Clin Rheumatol* 2008; 27: 327-32.
13. HOFFMAN DL, DUKES EM: The health status burden of people with fibromyalgia: a review of studies that assessed health status with the SF-36 or the SF-12. *Int J Clin Pract* 2008; 62: 115-26.
14. LOZA E, ABÁSOLO L, JOVER JA, CARMONA L, EPISER STUDY GROUP: Burden of disease across chronic diseases: a health survey that measured prevalence, function, and quality of life. *J Rheumatol* 2008; 35: 159-65.
15. DWORKIN RH, TURK DC, WYRWICH KW *et al*.: Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008: 9: 105-21.
16. BASCH E: Beyond the FDA PRO guidance: steps toward integrating meaningful patient-reported outcomes into regulatory trials and US drug labels. *Value Health* 2012; 15: 401-3.
17. TESTA MA, SIMONSON DC: Assessment of quality-of-life outcomes. *N Engl J Med* 1996; 28: 835-40.
18. SALAFFI F, SARZI-PUTTINI P, CIAPETTI A, ATZENI F: Assessment instruments for patients with fibromyalgia: properties, applications and interpretation. *Clin Exp Rheumatol* 2009; 27 (Suppl. 56): S92-105.
19. BURCKHARDT CS, CLARK SR, BENNETT RM: The Fibromyalgia Impact Questionnaire: development and validation. *J Rheumatol* 1991; 18: 728-33.
20. BENNETT RM, FRIEND R, JONES KD, WARD R, HAN BK, ROSS RL: The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties. *Arthritis Res Ther* 2009; 11: R120.
21. DUNKL PR, TAYLOR AG, MCCONNELL GG, ALFANO AP, CONAWAY MR: Responsiveness of fibromyalgia clinical trial outcome measures. *J Rheumatol* 2000; 27: 2683-91.
22. BENNETT R: The Fibromyalgia Impact Questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. *Clin Exp Rheumatol* 2005; 23 (Suppl. 39): S154-S162.
23. BOOMERSHINE CS: A comprehensive evaluation of standardized assessment tools in the diagnosis of fibromyalgia and in the assessment of fibromyalgia severity. *Pain Res Treat* 2012; 2012: 653-714.
24. SRIFI N, BAHIRI R, ROSTOM S, BENDED-DOUCHE I, LAZREK N, HAJJAJ-HASSOUNI N: The validity and reliability of the Moroccan version of the Revised Fibromyalgia Impact Questionnaire. *Rheumatol Int* 2013; 33: 179-83.
25. EDIZ L, HIZ O, TOPRAK M, TEKEOGLU I, ERCAN S: The validity and reliability of the Turkish version of the Revised Fibromyalgia Impact Questionnaire. *Clin Rheumatol* 2011; 30: 339-46.
26. WOLFE F, SMYTHE HA, YUNUS MB *et al*.: The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the Multicenter Criteria Committee. *Arthritis Rheum* 1990; 33: 160-72.
27. SALAFFI F, SARZI-PUTTINI P, GIROLIMETTI R, GASPARINI S, ATZENI F, GRASSI W: Development and validation of the self-administered Fibromyalgia Assessment Status: a disease-specific composite measure for evaluating treatment effect. *Arthritis Res Ther* 2009; 11: R125.
28. IANNUCCELLI C, SARZI-PUTTINI P, ATZENI F *et al*.: Psychometric properties of the Fibromyalgia Assessment Status (FAS) index: a national web-based study of fibromyalgia. *Clin Exp Rheumatol* 2011; 29 (Suppl. 69): S49-54.
29. GUILLEMIN F, BOMBARDIER C, BEATON D: Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417-32.
30. SALAFFI F, CIAPETTI A, SARZI-PUTTINI P *et al*.: Preliminary identification of key clinical domains for outcome evaluation in fibromyalgia using the Delphi method: the Italian experience. *Reumatismo* 2012; 19: 27-34.
31. STREINER DL, NORMAN GR: Health measurement scales. A practical guide to their development and use. 2nd ed., Oxford, Oxford University Press, 1995: 163-80.
32. HORN JL: A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965; 30: 179-85.
33. JÖRESKOG KG: Structural Equation Modeling with Ordinal Variables using LISREL. Scientific Software International: Lincolnwood, IL 2005. Available at: http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf. Accessed July 20, 2012.
34. GUADAGNOLI E, VELICER WF: Relation of sample size to the stability of component patterns. *Psychol Bull* 1988; 103: 265-75.
35. HU LT, BENTLER PM: Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999; 6: 1-55.
36. LINACRE JM: A user's guide to Winsteps-Ministep: Rasch-model computer programs. Program manual 3.68.0. Chicago, IL: WINSPTEPS.com; 2009. Available at: http//winsteps.com/a/winsteps.htm. Accessed February 15, 2010.
37. BOND TG, FOX CM: Applying the Rasch model: fundamental measurement in the human sciences. Mahwah, Lawrence Erlbaum Associates, 2001: 158-86.
38. FRANCHIGNONI F, GIORDANO A, SARTORIO F, VERCELLI S, PASCARIELLO B, FERRIERO G: Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch validation study. *Arch Phys Med Rehabil* 2010; 91: 1370-7.
39. LINACRE JM: Optimizing rating scale category effectiveness. *J Appl Meas* 2002; 3: 85-106.
40. FROST MH, REEVE BB, LIEPA AM, STAUFFER JW, HAYS RD, MAYO/FDA PATIENT-REPORTED OUTCOMES CONSENSUS MEETING GROUP: What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007; 10 (Suppl. 2): S94-105.
41. MEASE P, ARNOLD LM, CHOY EH *et al*.: OMERACT Fibromyalgia Working Group. Fibromyalgia syndrome module at OMERACT 9: domain construct. *J Rheumatol* 2009; 36: 2318-29.
42. VELICER WF, FAVA JL: Effects of variable and subject sampling on factor pattern recovery. *Psychol Methods* 1998; 3: 231-51.
43. FRANCHIGNONI F, GIORDANO A, FERRIERO G, ORLANDINI D, AMORESANO A, PERUCCA L: Measuring mobility in people with lower limb amputation: Rasch analysis of the mobility section of the prosthesis evaluation questionnaire. *J Rehabil Med* 2007; 39: 138-44.
44. KHADKA J, GOTHWAL VK, MCALINDEN C, LAMOUREUX EL, PESUDOVS K: The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes* 2012, 10: 80.
45. SARZI-PUTTINI P, ATZENI F, DI FRANCO M *et al*.: Dysfunctional syndromes and fibromyalgia: a 2012 critical digest. *Clin Exp Rheumatol* 2012; 30 (Suppl. 74): S143-51.
46. ALCIATI A, SARZI-PUTTINI P, BATTICCIOTTO A *et al*.: Overactive lifestyle in patients with fibromyalgia as a core feature of bipolar spectrum disorder. *Clin Exp Rheumatol* 2012; 30 (Suppl. 74): S122-8.
47. DÖNMEZ S, PAMUK ÖN, ÜMIT EG, TOP MŞ: Autoimmune rheumatic disease associated symptoms in fibromyalgia patients and their influence on anxiety, depression and somatisation: a comparative study. *Clin Exp Rheumatol* 2012; 30 (Suppl. 74): S65-9.