Evaluation of Patients With Fibromyalgia. Comparative Study of 4 Spanish Versions of the Fibromyalgia Impact Questionnaire (FIQ)

Joaquim Jordi Esteve Vives and Enrique Batlle Gualda
Sección de Reumatología, Hospital General Universitario de Alicante, Alicante, Spain.

Objective: To describe and compare the process of transcultural adaptation in the 4 validated Spanish versions of the Fibromyalgia Impact Questionnaire (FIQ).

Material and methods: a) Questionnaires. The first version (FIQ1) appeared as a doctoral thesis in 1988; the second (FIQ2) was published in a Spanish psychology journal; the third (FIQ3) was published in an English language journal; and the last (FIQ4) appeared in November 2004 in the Revista Española de Reumatología. b) Method. In each of the versions the following were assessed: 1) the semantic equivalence with respect to the original FIQ, 2) each version’s degree of development following a standardized method based on the GRAQol Index (GI), and 3) the impact of publication.

Results: The FIQ4 showed a greater semantic equivalency. The degree of development shown by the GI produced the following results: FIQ1, 56%; FIQ2, 50%; FIQ3, 75%; FIQ4, 31%. Only the FIQ3 results were published in MedLine-indexed journals.

Conclusion: The Spanish FIQ3 version presents a greater degree of development and an acceptable semantic equivalency with respect to the original, and has achieved a greater impact.

Key words: Fibromyalgia. Quality of life. Questionnaires. Translations. Validation studies.

Evaluación de pacientes con fibromialgia. Estudio comparativo de 4 versiones españolas del Fibromyalgia Impact Questionnaire (FIQ)

Objetivo: Describir y comparar el proceso de adaptación transcultural de las 4 versiones españolas validadas del Fibromyalgia Impact Questionnaire (FIQ).

Correspondence: Dr. J.J. Esteve Vives. Sección de Reumatología, Hospital General Universitario de Alicante, Avda. Pintor Baeza, s/n. 03010 Alicante. España. E-mail: ximo.esteve@ono.com

Manuscript received May 6, 2006; accepted for publication September 15, 2006.

Introduction

Fibromyalgia is a multidimensional illness characterized by generalized chronic pain and the frequent coexistence of fatigue, sleep disorders, anxiety, depression, and multiple other affections in a variable degree. The scarcity and inconsistency of the exploratory findings have caused that the evaluation of fibromyalgia be based solely on subjective measures. In this sense attention must be called to the usage of health questionnaires such as SF-36 and, above all, the Fibromyalgia Impact Questionnaire (FIQ) that by its characteristic brevity, ease of application, and multidimensionality has reached great acceptance on a global scale. Proof of this is the fact that it has been translated and adapted to multiple languages such as German, Korean, French, Turkish, Italian, Hebrew, or Swedish. In Spain we have 4 versions of the FIQ that...
have followed, all of them, an elaborate process of translation and adaptation. Even more, in a review of the Spanish literature we still find 2 more versions of the FIQ adapted to evaluate the response to specific treatments in patients with fibromyalgia.\textsuperscript{9,10} Faced with so many versions of the same instrument and with the goal of clarifying this situation, this study was planned with the objective of describing and comparing the process of transcultural adaptation of each one of the 4 validated Spanish versions of the FIQ.

**Material and Methods**

**Questionnaires**

The original FIQ, published by Burckhardt et al\textsuperscript{12} in 1991, is an autoapplied questionnaire of 10 items. The first item, named scale of physical function, is formed by another 10 items, al of Likert scale type response with 4 levels (0, always able to do; to 3, never able to do). On item 2 the patient must point out the number of days in which he or she felt well during the past week. Items 3 and 4 refer to work related activities of the patient: number of working days lost during the past week and degree of difficulty to work respectively. The 6 remaining items, the same as item 4, are scored using Visual Analog Scales (VAS) of 100 mm and their content evaluates pain, fatigue, morning fatigue, stiffness, anxiety, and depression. The way to obtain the final score consists of standardizing all of the items on a 1 to 10 scale and adding, afterward, the scores; the scoring of the questionnaire can oscillate between 0 and 80 or between 0 or 100, representing, in both cases, a worse health state with higher punctuations. The first Spanish version of the FIQ is the work of B. González et al (FIQ\textsubscript{1}), adapted to evaluate the response to specific treatments in patients with fibromyalgia. \textsuperscript{15} The psychometric characteristics came from 41 women from the Asociación Catalana de Afectados de Fibromialgia and the scoring range employed was 0 to 100.

A particularity of this questionnaire was the work of M. de Gracia et al and it was published in a Spanish psychology journal.\textsuperscript{16} The psychometric characteristics were evaluated in a simple of 73 patients with FM (71 women and 2 men) referred from the outpatient consult of rheumatology. The range of scoring of the questionnaire was 0 to 10 because the authors opted to find the arithmetic mean of the group of items of the questionnaire. The second version, FIQ\textsubscript{2}, presented initially as a doctoral thesis\textsuperscript{17} in 1998 and communicated initially in the annual meeting of the American College of Rheumatology (ACR)\textsuperscript{18} in 1999. The psychometric characteristics were evaluated in a simple of 41 women with fibromyalgia that CAME from different centers for functional rehabilitation and psychology clinics. The scoring range employed was 0 to 10, the same as in FIQ\textsubscript{1}. The third version, FIQ\textsubscript{3}, developed by J. Rivera et al was published in English in a rheumatology journal.\textsuperscript{19} The psychometric characteristics were evaluated in a simple of 102 women with fibromyalgia that CAME from a rheumatology outpatient and the scoring range employed was from 0 to 80. The fourth version was developed by S. Monterde et al and published in the Revista Española de Reumatología.\textsuperscript{17} The psychometric characteristics came from 41 women of the Confederação de Afectados de Fibromialgia and the scoring range employed was 0 to 100.

**Comparative Analysis of the 4 Spanish Versions of the FIQ**

The authors were initially contacted to obtain a copy of each version to study (A) the semantic equivalent with respect to the original FIQ, (B) the level of development in each one of them, and (C) the impact of their publications.

**A. Semantic Equivalence**

A professional bilingual translator, originally from the United Kingdom, did a blinded retrotranslation of each one of the Spanish versions of the FIQ. After this, she compared them to the original version in English according to a standardized criteria employed beforehand,\textsuperscript{18} that consisted in classifying the items in 3 groups according to their level of agreement: items A (satisfactory agreement), the formulation and sense are equal to the original item; items B (quite satisfactory agreement), the formulation is not the same and there can be some discordant words but the item captures the sense of the original; items C (no agreement) the formulation and the sense of the item are different from the original.

**B. Level of Development**

A standardized method based on the GRAQoL index (GI) was carried out,\textsuperscript{20} according to which 8 different aspects or criteria of the transcultural process adaptation of each of the 4 Spanish versions of the FIQ were evaluated; each aspect was scored from 0 to 2 and the result expressed in percentage points. It was considered that a GI between 50% and 70% had an acceptable level and if <50% the development level was poor. The 4 versions were independently evaluated by the 2 authors of this study and the discrepancies were solved by consensus afterwards. The 8 criteria evaluated were the following:

1. Translations and retrotranslations. The process of translation to Spanish was evaluated, especially considering if any retrotranslations to English had been done.

2. Piloting. Piloting with patients to detect transcultural differences that could invalidate some aspect of the questionnaire was evaluated.
3. Structural validity. A structural study was carried out using factor analysis with the scores of the questionnaires.

4. Convergence–discriminant validity. Scores of the questionnaire were compared to other instruments that supposedly measure the same concept (convergence validity) and others that measure other concepts (discriminant validity). That way the evaluation of what measurements and instruments were employed to compare the other 4 Spanish versions of the FIQ.

5. Sensibility of the questionnaires in different populations.

6. Analysis of the internal consistency. It was analyzed through the Cronbach coefficient.

7. Test–retest reproducibility analysis. Done using the coefficient or agreement determination between the baseline punctuations of the different items and those obtained after 1 or a few weeks. In a complementary fashion, these results were compared to the ones of the original version of the FIQ.

8. Sensibility to change. The presence of a sensibility to change analysis of the questionnaires alter some therapeutic intervention previously considered effective was evaluated.

C. Impact of Publication

A MedLine literature search using the key words “fiq” and “Spanish” or “fibromyalgia impact questionnaire” and “Spanish”. In a complimentary manner we contacted the authors of the 4 versions of FIQ and the literature databases “Spanish”. In a complimentary manner we contacted the authors of the 4 versions of FIQ and the literature databases of the Consejo Superior de Investigaciones Científicas (CSIC) were consulted using the key words “fibromyalgia impact questionnaire” or “fiq” or “fibromialgia”, and “cuestionario”.

Results

A. Semantic Equivalence

FIQ4 had the maximum agreement with the original FIQ, (Table 1) because 16 items (84%) were classified as “A,” and none were classified as “no agreement” with the original version. The least agreement was found with FIQ1, while versions FIQ2 and FIQ3 were situated in an intermediate position. For example, items A: “Ir a la compra” (FIQ1), “Ir de compras” (FIQ2), “Hacer la compra” (FIQ3), or “Ir a comprar” (FIQ4) were considered equivalent to the original “Do shopping.”

~Items C: “¿Se ha encontrado rigido? No/Si” (FIQ1) was not considered concordant with the original “How bad has your stiffness been? No stiffness/Very stiff.” The item “Utilizar transporte público” (FIQ3) was not considered concordant with the original “Drive a car” although, in this case, the lack of agreement was deliberately looked for by the authors to prove that the great majority of their patients did not drive a car.

B. Level of Development. GRAQoL Index

The evaluation of the GI by the 2 authors of this study was identical in the case of criteria 2, 3, 5, 6, and 7; in the remaining 3 criteria there were some discrepancies that were solved by consensus without that qualitatively affecting the results that are summarized on table 2 and commented below.

Criteria 1. Only the authors of FIQ2 and FIQ4 followed an organized process of translation/retrotranslation and final consensus. The rest did various translations of the original, and their differences were solved by consensus and were finally reviewed by a specialized translator.

Criteria 2. The authors of FIQ1 and FIQ3 did large pilot studies with their initial versions and this permitted showing that some subitems were not relevant in the Spanish population of patients with fibromyalgia, fundamentally “Drive a car” and “Do yardwork.”

Criteria 3. The structural validity was only evaluated by the authors of FIQ15 who found two dimensions: one grouped the intensity of pain, sadness, stiffness, and in a lesser way limitations for labor; the other one grouped the items referred to anxiety, physical function, and fatigue.

Criteria 4. The evaluation of the convergent–discriminant validity was done in an unequal manner by the authors of the 4 adaptations: González et al13,14 compared the punctuations of the FIQ1 with variables such as age, intensity of the symptoms of fibromyalgia, the ACR tender point count, and the pain threshold evaluated using a pain meter of 9 predetermined points. The agreement was low, though significant, between the global store of the FIQ1 and the number of tender points (r=0.29) or the pain threshold (r=0.37). De Gracia et al15 compared the store
of the FIQ2 with the ones from the psychopathologic symptoms measurement questionnaire SCL-90-R and with intensity variables of various symptoms of fibromyalgia, without doing any comparison with the ACR tender points. The VAS of pain today 286 Reumatol Clin. 2006;2(6):283-8 SCL-90-R and with intensity variables of various

The formula:

psychopathologic symptoms measurement questionnaire 1, has been done but is insufficient; 2, has been done. The GI is obtained using one of the enumerated criteria. 0 indicates has not been done or is unknown; *A score of 0 to 2 is assigned according to the degree of compliance of each

The convergent-discriminant-

Validity of the FIQ*

The test-retest reproducibility was studied by repeating the administration of the questionnaire after 1 week. Their results were similar in versions FIQ and FIQ3 (Table 3), and the study of FIQ 4 included a scarce number of patients and did not attain statistical significance in 3 of the 10 items of the questionnaire.

The sensibility to change was satisfactorily evaluated by the authors of FIQ3, through a parallel clinical trial that compared the efficacy of one program of aerobic fitness evaluated. The internal consistence was not evaluated for FIQ4.

Criteria 6. The internal trustworthiness showed similar results in versions FIQ1 and FIQ3, with an alpha Cronbach coefficient in both of 0.82 for the group of the 10 principal items. Of the FIQ2 version we evaluated the internal consistence of 8 out of the 10 items (=0.93), after

0.61), in contrast with the correlation between the anxiety scales of the FIQ and the SCL-90 which were unexpectedly low (r=0.05, P=ns). Rivera et al10 analyzed the correlations between FIQ3 and the functional capacity questionnaire HAQ, in its reduced version specific for fibromyalgia (FHAQ, a spanish version of the global health questionnaire SF-36, a VAS for pain, the ACR tender point count, and the SCL-90-R questionnaire that evaluates psychological symptoms). Of notice was the low but significant correlation between the anxiety and depression items of FIQ4 and the mental health function scales of both questionnaires.

Criteria 5. The sensibility in different populations was not evaluated in any case.

Criteria 6. The internal trustworthiness showed similar results in versions FIQ1 and FIQ3, with an alpha Cronbach coefficient in both of 0.82 for the group of the 10 principal items. Of the FIQ2 version we evaluated the internal consistence of 8 out of the 10 items (=0.93), after excluding the analysis of the items referring to days in which the patient felt good or lost working days. Internal consistence was not evaluated for FIQ4.

Criteria 7. The test-retest reproducibility was studied by repeating the administration of the questionnaire after 1 week. Their results were similar in versions FIQ and FIQ3 (Table 3), and the study of FIQ 4 included a scarce number of patients and did not attain statistical significance in 3 of the 10 items of the questionnaire.

Criteria 8. The sensibility to change was satisfactorily evaluated by the authors of FIQ3, through a parallel clinical trial that compared the efficacy of one program of aerobic fitness to a psychological intervention of the
day. Internal consistence was not evaluated for FIQ4.

Criteria 6. The internal trustworthiness showed similar results in versions FIQ1 and FIQ3, with an alpha Cronbach coefficient in both of 0.82 for the group of the 10 principal items. Of the FIQ2 version we evaluated the internal consistence of 8 out of the 10 items (=0.93), after excluding the analysis of the items referring to days in which the patient felt good or lost working days. Internal consistence was not evaluated for FIQ4.

Criteria 7. The test-retest reproducibility was studied by repeating the administration of the questionnaire after 1 week. Their results were similar in versions FIQ and FIQ3 (Table 3), and the study of FIQ 4 included a scarce number of patients and did not attain statistical significance in 3 of the 10 items of the questionnaire.

Criteria 8. The sensibility to change was satisfactorily evaluated by the authors of FIQ3, through a parallel clinical trial that compared the efficacy of one program of aerobic physical fitness to a psychological intervention of the conduct-cognitive type. Their results showed that the score (±SD) of FIQ3 (0 to 80) improved significantly from 52.0 (±11.5) to 40.8 (±13.7) after the compliance of the physical fitness program, as well as the variables of physical fitness evaluated. In all, the GRAQoL index showed a good level of development, the FIQ3 an acceptable level of development, the FIQ2 a poor level of development, and the FIQ4 a poor level of development (Table 2).

C. Impact of Publication

Only FIQ3 has been published in a MedLine indexed journal. FIQ2 and FIQ3 have been published in

TABLA 2. GRAQoL Index (GI) of the 4 Spanish Versions of the FIQ*

<table>
<thead>
<tr>
<th>FIQ1</th>
<th>FIQ2</th>
<th>FIQ3</th>
<th>FIQ4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Translation and retrotranslation</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2. Pilot study of the adaptation</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3. Structural validity</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4. Convergent-discriminant-validity</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5. Sensibility in different populations</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. Internal trustworthiness</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7. Test-retest reproducibility</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8. Sensibility to change</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GRAQoL index:</td>
<td>56%</td>
<td>50%</td>
<td>75%</td>
</tr>
</tbody>
</table>

* A score of 0 to 2 is assigned according to the degree of compliance of each one of the enumerated criteria. 0 indicates has not been done or is unknown; 1, has been done but is insufficient; 2, has been done. The GI is obtained using the following formula:

GI = \frac{\text{Sum of Scores}}{\text{Maximum Possible Score}} \times 100

TABLE 3. Test-Retest Reproducibility of the Spanish Versions of the FIQ and Their Comparison to the Original Version*

<table>
<thead>
<tr>
<th>FIQ1</th>
<th>FIQ2</th>
<th>FIQ3</th>
<th>FIQ4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r (n)</td>
<td>rS (n)</td>
<td>r (n)</td>
<td>rS (n)</td>
</tr>
<tr>
<td>Item 1</td>
<td>0.88 (1)</td>
<td>0.79 (2)</td>
<td>0.56 (3)</td>
</tr>
<tr>
<td>Item 2</td>
<td>0.74 (5)</td>
<td>0.68 (6)</td>
<td>0.32 (7)</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.71 (9)</td>
<td>0.83 (10)</td>
<td>0.67 (11)</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.82 (13)</td>
<td>0.76 (14)</td>
<td>0.53 (15)</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.7 (17)</td>
<td>0.74 (18)</td>
<td>0.32 (19)</td>
</tr>
<tr>
<td>Item 6</td>
<td>0.66 (21)</td>
<td>0.6 (22)</td>
<td>0.32 (23)</td>
</tr>
<tr>
<td>Item 7</td>
<td>0.64 (25)</td>
<td>0.6 (26)</td>
<td>0.32 (27)</td>
</tr>
<tr>
<td>Item 8</td>
<td>0.63 (29)</td>
<td>0.6 (30)</td>
<td>0.32 (31)</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.73 (33)</td>
<td>0.7 (34)</td>
<td>0.32 (35)</td>
</tr>
<tr>
<td>Item 10</td>
<td>0.91 (37)</td>
<td>0.7 (38)</td>
<td>0.32 (39)</td>
</tr>
</tbody>
</table>

* The test-retest reproducibility is unknown for FIQ4 and some partial aspects of FIQ. FIQ indicates Original FIQ; n, number of patients included in each study; r, Pearson correlation coefficient; rS, Spearman correlation coefficient.

The method for evaluating the test-retest reproducibility of FIQ was different in each case. This indicates that the correlations of this item did not reach statistical significance (P=ns).
The transcultural adaptation of a health measurement instrument has turned, in the past few years, into a relatively standardized process and its steps can be summarized in a schematic form in the 11 criteria of the GRAQoL index, of which we only used the 8 applicable to the type of instrument represented by the FIQ in this study (Table 2). The FIQ, developed by González et al.13,14 was, chronologically, the first Spanish version of the FIQ; although it had the inconvenience of a scarce diffusion. The GRAQoL index showed an acceptable development level, but the semantic analysis showed an excessively poor agreement with respect to the original FIQ, not explained by the transcultural adaptation process. Manuel de Gracia et al developed the first version of FIQ to be published in a journal (FIQ1).15 That study gave an interesting factor analysis, though its results were different from those found in the original FIQ by Burckhardt et al.90-R. The most notable aspect of FIQ4, Developer by Monterde et al.9 was, semantically, the first Spanish version of the FIQ, although it had the inconvenience of a scarce diffusion. The GRAQoL index showed an acceptable development level, but the semantic analysis showed an excessively poor agreement with respect to the original FIQ, not explained by the transcultural adaptation process. The semantic equivalent of this version with respect to the original was elevated and its level of development was relatively acceptable, but no piloting with patients was done. The number of patients included in the evaluation of the psychometric characteristics was scarce (41) which motivated the unexpectedly low correlation between the anxiety scales and of the FIQ and the SLSA 90-R. The most notable aspect of FIQ4, Developer by Monterde et al.9 was its elevated agreement level with respect to the original, with an acceptable level of semantic agreement with respect to the original. Its level of development, nonetheless, was relatively poor and the number of patients, scarce (41). This fact probably motivated a low test-retest reproducibility of items 2 and 4 of this version (Table 3). Lastly, the FIQ3 version of Rivera et al.16 was the one with a highest level of development, with a semantic agreement considered acceptable with respect to the original FIQ and a publication in a journal with a larger impact, being the only one indexed in MedLine. The study of convergent-discriminant validity17 and the sensitivity to change18 were especially interesting. Once the comparative study was finalized it is important to point out that the recent publication of an updated version of the original FIQ5 that has definitely established the scoring system of 0 to 100, has incorporated the consideration of the domestic work in items 3 and 4, and has substituted the VAS of the last 7 items for graded scales (0 to 10) and has added a sub item (“climb stairs”) to the physical function scale. In this sense, our work team has proposed the elaboration of an updated version of the FIQ5 that parting from FIQ3, takes into consideration some important aspects of the other Spanish versions of the FIQ and the recent updated original version.

Acknowledgment
To Drs Cayetano Alegre, Manuel de Gracia, Javier Rivera, and Isabel Salvat for their collaboration in providing all of the information on their versions of the FIQ, that was requested. To the medical team of the Hospital de Sant Vicent ( Alicante) for their collaboration in lending the microphotographic version of the doctoral thesis of FIQ1 which is in us out of print. To Jessica Gorlin, our bilingual translator and friend.

References


