Concise Report

Validation of a Dutch translation of the fibromyalgia impact questionnaire

T. R. Zijlstra¹, E. Taal², M. A. F. J. van de Laar¹,² and J. J. Rasker²

Objectives. To validate a Dutch translation of the fibromyalgia impact questionnaire (FIQ).

Materials and methods. Data were taken from two randomized clinical trials on Spa treatment and venlafaxine in fibromyalgia (FM). Participants completed the Dutch FIQ and a set of validated questionnaires for general health (RAND-36), depression (Beck depression inventory, BDI), pain (McGill pain questionnaire, MPQ) and fatigue (checklist individual strength, CIS). Internal consistency within the FIQ item ‘physical functioning’ was studied using Cronbach’s α. Test–retest reliability was studied with intra-class-correlation (ICC) in a subsample of 76 control subjects over a 3 month period without specific intervention. Construct validity was evaluated by correlating the FIQ to other questionnaires. Sensitivity to change was studied using standardized response means (SRM).

Results. The study sample consisted of 213 women and 11 men (mean age 47 yrs, mean disease duration 11 yrs). Cronbach’s α for the item ‘physical functioning’ was 0.91, indicating high internal consistency. Test–retest reliability was acceptable, with ICC ranging from 0.45 for ‘morning tiredness’ to 0.71 for ‘physical function’. FIQ correlated significantly with the RAND-36, with Spearman’s ρ ranging from −0.60 to −0.70 for items measuring the same concept. Similar patterns of correlation were seen with MPQ, BDI and CIS. Sensitivity to change was sufficient, with SRM after Spa treatment ranging from 0.3 for ‘work days missed’ to 0.9 for ‘days felt good’. Similar SRM were found in the venlafaxine trial for patients reporting general improvement.

Conclusion. The Dutch FIQ is a valid instrument for measuring health status in FM, showing sufficient reliability, construct validity and responsiveness.

KEY WORDS: Fibromyalgia impact questionnaire, Fibromyalgia, Dutch version, Validation, Health assessment.

The fibromyalgia impact questionnaire (FIQ) is a questionnaire to assess the health status of women with fibromyalgia (FM) syndrome [1]. It has been translated into several languages, and has been recommended as a primary end-point in FM clinical trials [2–10]. In order to join this international development, we set out to make a Dutch version of the FIQ and evaluate its psychometric properties.

Materials and methods

Patients

Data were used from two studies among patients aged 18–65 yrs with primary FM, according to 1990 American College of Rheumatology (ACR) criteria [11].

Study 1. This was a randomized controlled trial of Spa treatment (n = 58) vs treatment as usual (n = 76). For details see Zijlstra et al. [12].

Study 2. This was a 6 week, double blind, randomized trial of venlafaxine (n = 45) vs placebo (n = 45). Subjects were recruited from four out-patient rheumatology clinics, and through an advertisement in the Dutch FM Patient Association Magazine. Exclusion criteria were: severe depression needing drug treatment, use of antidepressive drugs in the past 6 weeks, previous use of venlafaxine, (chance of) pregnancy and medication or comorbidity interfering with venlafaxine. This study has only been published as an abstract [13].

Measures

FIQ. The FIQ consists of 10 items. The first item contains 10 questions on activities of daily living, each of which are scored in a Likert format from 0 (always able to do) to 3 (never able to do). The scores are added and divided by the number of valid scores to yield one score for physical functioning. Item 2 is the number of days (0–7) felt good during the past week. Item 3 asks for the number of days off work during the past week (0–5). Items 4–10 (ability to do job, pain, fatigue, morning tiredness, stiffness, anxiety and depression) are measured by 100 mm visual analogue scales. The scores of each item are standardized on a scale ranging from 0–10 with higher scores indicating greater impairment.

The FIQ was translated from English into Dutch by one of the investigators (T.R.Z.). Two adaptations were made. In item 3,
the maximum number of work days missed was reduced to five, in accordance with a normal working week in the Netherlands. In the 4th item, ‘When you did go to work,’ was left out and the word ‘job’ was translated with ‘werkzaamheden’ (meaning ‘activities’). In this way the question also applies to those without a professional job, who make up a considerable part of the Dutch FM population. We had our Dutch version back-translated by a Dutch teacher of English. When comparing the original and the back-translation, we found some differences, but these stem from our cultural adaptations in items 3 and 4.

**RAND-36.** This is a validated Dutch translation of the Short Form-36 (SF-36), a generic measure of self-reported health status, also used in FM patients [14–18]. It covers eight domains of functioning and well-being. Domain scores are standardised on a 0–100 scale with higher scores indicating better status.

**Beck depression inventory (BDI).** Depression was measured with a Dutch version of the BDI [19, 20]. The score ranges from 0 to 63.

**McGill pain questionnaire (MPQ).** The Dutch language version of the MPQ comprises a set of 20 groups of three or four words describing several characteristics of pain [21, 22]. Within each group there is an increase in severity, reflected by an increase in score. The scores are summed to yield a total pain rating index (PRI-T, range 0–63).

**Checklist individual strength (CIS).** The CIS was used to measure fatigue [23]. The questionnaire consists of 11 statements, eight concerning subjective feelings of fatigue, and three concerning activity. Each item is scored on a 7-point Likert scale from 1 (yes, correct) to 7 (no, incorrect). Thus, the scores range from 8–56 for subjective feelings and 3–21 for activity.

**General improvement.** In study 2, at 6 weeks, subjects were asked if they had experienced general improvement during the past weeks (yes/no).

**Tender point assessment.** A single observer (T.R.Z.) examined each patient by manual palpation of the 18 body sites defined in the ACR criteria for FM [12]. The tender point score (TPS) was the number of sites on which the patient stated that palpation was painful (range 0–18). For the graded tender point score (GTPS) each tender point was scored by the observer as 0 (no pain), 1 (mild pain, no grimace), 2 (spontaneous verbal reaction to pain and grimace), 3 (severe pain with withdrawal) and the sum of 18 points was recorded [24].

**Procedures**

**Study 1.** All subjects completed the questionnaires, and tender point examinations were performed at baseline and at 3, 6 and 12 months during the follow-up period of the RCT. Subjects in the intervention group were also assessed at 1 month, which was 1 week after the end of the Spa treatment.

**Study 2.** Assessments were performed at baseline and at 6 weeks.

**Statistics**

Internal consistency of the 10 questions in the first item of the FIQ, physical functioning, was studied using Cronbach’s $\alpha$.

Test–retest reliability was studied with intra-class-correlation (ICC) over a period of 3 months in a subsample of 76 control subjects from study 1.

To study convergent validity of the FIQ, Spearman’s $\rho$ correlation coefficients were calculated with other questionnaires and tests.

Responsiveness was assessed in two ways. Pre- and post-test (1 month) results of the FIQ were compared in a subsample of 58 subjects receiving Spa treatment in study 1. In study 2, responsiveness was analysed from the patients’ perspective by comparing pre- and post-test results of the FIQ in patients who had experienced general improvement and in patients who did not experience general improvement irrespective of treatment with venlafaxine or placebo. Significance of changes in FIQ scores from pre- to post-test was tested by Wilcoxon signed rank tests. Standardized response means (SRM) were calculated as the ratio of the mean change between baseline and 6 weeks to the standard deviation (S.D.) of that change. The 95% confidence intervals (CI) for the SRM were calculated under the assumption that change scores followed a normal distribution and therefore the distribution of the SRM could be approximated by a normal distribution, with a mean of 0 and a s.d. of 1 divided by the square root of the sample size [25].

**Results**

The study sample consisted of 213 women and 11 men. Mean age was 47 yrs (range: 22–68, s.d. = 9.3) and mean disease duration was 11 yrs (range: 0–42, s.d. = 8.2). The percentage of missing values was below 3 for most items of the FIQ. The question ‘Were you able to drive a car?’ was not applicable to 32 patients (14%) without a driver’s license; and item 3 about ‘work days missed’ was only applicable to 91 patients (41%) because the other patients were unemployed.

Internal consistency of the ‘physical functioning’ scale was excellent (Cronbach’s $\alpha$ = 0.91).

The means and s.d. at baseline and test–retest results are presented in Table 1. ICC coefficients of test–retest reliability varied from 0.45 to 0.71.

In general, FIQ items correlated best with corresponding items of the RAND-36 (Table 2). Thus, FIQ-physical function correlated best with RAND-physical function, FIQ-pain with RAND-pain, FIQ-fatigue and -morning tiredness with RAND-vitality, and FIQ-anxiety and -depression with RAND-role emotional and -mental health. Correlations between the FIQ and other outcome measures showed the same pattern (supplementary Table 3). FIQ-anxiety and -depression correlated best with BDI (0.52 and 0.54), FIQ-pain with MPQ (0.44), and FIQ-fatigue and -morning tiredness with CIS-subjective (0.61 and 0.53). All FIQ items correlated poorly with tender point scores (all $R < 0.23$).

One week after Spa treatment the 58 subjects reported statistically significant improvement ($P \leq 0.01$) on all FIQ items, with the exception of ‘work days missed’. However, this scale was only applicable for 22 patients, and baseline scores were already very good (mean = 0.8). SRMs ranged from 0.3 for ‘work days missed’ and for ‘depression’ to 0.9 for ‘days felt good’ (supplementary Table 4).

The FIQ scores also showed to be sensitive for changes from the patient’s perspective (supplementary Table 5). Patients who experienced general improvement after 6 weeks had significantly improved scores ($P \leq 0.01$) for most FIQ items, while patients who did not experience general improvement showed no improved FIQ scores. FIQ items ‘work days missed’ (only applicable for 11 patients, and a very good mean baseline score of 0.4) and ‘anxiety’ showed not to be responsive for perceived general improvement. SRMs for patients who perceived to be improved for the other FIQ items ranged from 0.6 for ‘fatigue’ and ‘depression’ to 1.5 for ‘days felt good’. SRMs for patients who perceived not to be improved ranged from $-0.4$ to 0.2.
Depression / Anxiety / Stiffness / Morning tiredness / Fatigue

<table>
<thead>
<tr>
<th>FIQ-item</th>
<th>Study 1 (n=134)</th>
<th>Study 2 (n=90)</th>
<th>All patients (n=224)</th>
<th>Test-retest reliabilitya (n=76)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Physical function</td>
<td>4.5</td>
<td>1.6</td>
<td>4.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Days felt good</td>
<td>6.9</td>
<td>2.6</td>
<td>7.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Work days missedb</td>
<td>0.8</td>
<td>2.4</td>
<td>2.8</td>
<td>3.9</td>
</tr>
<tr>
<td>Job ability</td>
<td>5.7</td>
<td>2.1</td>
<td>5.9</td>
<td>2.3</td>
</tr>
<tr>
<td>Pain</td>
<td>5.8</td>
<td>1.7</td>
<td>6.5</td>
<td>1.7</td>
</tr>
<tr>
<td>Fatigue</td>
<td>6.4</td>
<td>2.0</td>
<td>6.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Morning tiredness</td>
<td>6.4</td>
<td>2.1</td>
<td>7.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Stiffness</td>
<td>6.4</td>
<td>2.0</td>
<td>6.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.5</td>
<td>2.6</td>
<td>3.8</td>
<td>2.9</td>
</tr>
<tr>
<td>Depression</td>
<td>2.7</td>
<td>2.4</td>
<td>2.9</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Table 1. Mean scores and standard deviations (s.d.) at baseline and test-retest reliability of FIQ items

Table 2. Non-parametric Spearman’s ρ correlations between FIQ and RAND-36 for 224 subjects with fibromyalgia

<table>
<thead>
<tr>
<th>RAND-36 dimensions</th>
<th>Physical function</th>
<th>Role physical</th>
<th>Pain</th>
<th>General health</th>
<th>Social function</th>
<th>Role emotional</th>
<th>Mental health</th>
<th>Vitality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical function</td>
<td>-0.60***</td>
<td>-0.33***</td>
<td>-0.40***</td>
<td>-0.36***</td>
<td>-0.44***</td>
<td>-0.14*</td>
<td>-0.10</td>
<td>-0.32***</td>
</tr>
<tr>
<td>Days felt gooda</td>
<td>-0.41***</td>
<td>-0.29***</td>
<td>-0.45***</td>
<td>-0.26***</td>
<td>-0.30***</td>
<td>-0.22***</td>
<td>-0.18***</td>
<td>-0.44***</td>
</tr>
<tr>
<td>Work days missedb</td>
<td>-0.27***</td>
<td>-0.21*</td>
<td>-0.35***</td>
<td>-0.08</td>
<td>-0.16</td>
<td>-0.13</td>
<td>-0.09</td>
<td>-0.26*</td>
</tr>
<tr>
<td>Job ability</td>
<td>-0.47***</td>
<td>-0.44***</td>
<td>-0.64***</td>
<td>-0.31***</td>
<td>-0.41***</td>
<td>-0.27***</td>
<td>-0.23***</td>
<td>-0.40***</td>
</tr>
<tr>
<td>Pain</td>
<td>-0.39***</td>
<td>-0.38***</td>
<td>-0.68***</td>
<td>-0.28***</td>
<td>-0.44***</td>
<td>-0.20***</td>
<td>-0.17*</td>
<td>-0.40***</td>
</tr>
<tr>
<td>Fatigue</td>
<td>-0.39***</td>
<td>-0.37***</td>
<td>-0.46***</td>
<td>-0.31***</td>
<td>-0.50***</td>
<td>-0.26***</td>
<td>-0.24***</td>
<td>-0.64***</td>
</tr>
<tr>
<td>Morning tiredness</td>
<td>-0.28***</td>
<td>-0.30***</td>
<td>-0.38***</td>
<td>-0.33***</td>
<td>-0.33***</td>
<td>-0.27***</td>
<td>-0.22***</td>
<td>-0.52***</td>
</tr>
<tr>
<td>Stiffness</td>
<td>-0.30***</td>
<td>-0.36***</td>
<td>-0.51***</td>
<td>-0.30***</td>
<td>-0.32***</td>
<td>-0.20***</td>
<td>-0.08</td>
<td>-0.36***</td>
</tr>
<tr>
<td>Anxiety</td>
<td>-0.24***</td>
<td>-0.16*</td>
<td>-0.22***</td>
<td>-0.37***</td>
<td>-0.25***</td>
<td>-0.47***</td>
<td>-0.70***</td>
<td>-0.33***</td>
</tr>
<tr>
<td>Depression</td>
<td>-0.15*</td>
<td>-0.19**</td>
<td>-0.16*</td>
<td>-0.37***</td>
<td>-0.31***</td>
<td>-0.51***</td>
<td>-0.70***</td>
<td>-0.36***</td>
</tr>
</tbody>
</table>

*P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001.

Discussion

This study showed the Dutch FIQ to be a reliable, valid and responsive measure.

In translating a questionnaire, cultural adaptations are at least as important as the linguistic aspects. Our adaptation in item 3 (a maximum of five working days per week) seems logical and is in line with similar adaptations in the Swedish, Hebrew and German versions [2, 3, 5]. After conversion to a 0–10 score, results from countries with working weeks of varying length can be compared directly. However, other problems remain. Firstly, this item does not take into account if someone has a full-time or part-time job. Perhaps it would be better to convert the score to a proportion of the individual’s normal working week, instead of the full-time working week. Secondly, item 3 only applies to subjects in paid employment. Since the percentage of women in paid employment may vary strongly between countries, it may prove difficult to compare results from different countries. For this reason, we also adapted item 4, using ‘activities’ instead of ‘job’. We prefer not to restrict this item to paid employment, although the impact of FM on unpaid activities such as housework may not be exactly the same.

The internal consistency of item 1 proved excellent. Although Offenbaecher and Sarz-Puttini reported even higher values (0.94), others found values of 0.88 or less [1–9].

Test-retest reliability (ICC) over three months ranged from 0.45 (morning tiredness) to 0.71 (physical function). Although it is usually recommended that the reliability should exceed 0.70 in stable patients, we believe the reliability to be acceptable because of the rather long time interval between measurements. All other studies except the French and Korean found better test–retest reliability [1–9]. This is probably due to a much shorter test–retest interval of 7–10 days, as compared to 3 months in our study. In other studies, test–retest reliability is measured with Spearman’s or Pearson’s correlation coefficient. We have used ICC instead because it is more appropriate. Correlation is a measure of association, and repeated measurements may be correlated but systematically different.

The significant correlations of most FIQ items with other corresponding outcome measures suggest that the FIQ has sufficient convergent validity. Only item 3 (work days missed) did not correlate significantly, probably because of the small number of subjects who had a job and the highly skewed distribution of the data. In our study, FIQ-pain and other FIQ items hardly correlated with TPSs. This is in line with findings by Jacobs et al. [24], who found a weak correlation between TPSs and self-reported pain. They concluded that TPSs and self-reported pain represent different aspects of pain in FM. Four validation studies also showed weak correlations between FIQ items and TPSs [1, 4, 6, 7]. However, other studies showed much stronger correlations [2, 5, 8, 9].

According to Cohen [26], an effect size (which is comparable to SRM) of 0.2 indicates a small, 0.5 a moderate and 0.8 a large effect. Thus, our version of the FIQ showed good sensitivity to change after Spa treatment. More importantly, patients who felt they had improved during the drug trial scored better on most FIQ items, while patients who felt they had not improved showed no improvement on the FIQ either. We can conclude that the Dutch FIQ is a responsive measure. Sensitivity
to change has been examined for only a few other language versions of the FIQ. The American and Spanish FIQ have been shown to be responsive to perceived clinical improvement [7, 10].

In conclusion, the Dutch FIQ is a reliable, valid and responsive outcome measure.

Acknowledgements
This study was funded by the Dutch Arthritis Association, grant NR 97-1-303.

The authors have declared no conflicts of interest.

References