Minimal Clinically Important Difference in the Fibromyalgia Impact Questionnaire

ROBERT M. BENNETT, ANDREW G. BUSHMAKIN, JOSEPH C. CAPPELLERI, GERGANA ZLATEVA, and ALESIA B. SADOSKY

ABSTRACT. Objective. The Fibromyalgia Impact Questionnaire (FIQ) is a disease-specific composite instrument that measures the effect of problems experienced by patients with fibromyalgia (FM). Utilization of the FIQ in measuring changes due to interventions in FM requires derivation of a clinically meaningful change for that instrument. Analyses were conducted to estimate the minimal clinically important difference (MCID), and to propose FIQ severity categories.

Methods. Data from 3 similarly designed, 3-month placebo-controlled, clinical treatment trials of pregabalin 300, 450, and 600 mg/day in patients with FM were modeled to estimate the change in the mean FIQ total and stiffness items corresponding to each category on the Patient Global Impression of Change. FIQ severity categories were modeled and determined using established pain severity cutpoints as an anchor.

Results. A total of 2228 patients, mean age 49 years, 93% women, with a mean baseline FIQ total score of 62 were treated in the 3 studies. Estimated MCID on a given measure were similar across the studies. In a pooled analysis the estimated MCID (95% confidence interval) was 14% (13; 15) and for FIQ stiffness it was 13% (12; 14). In the severity analysis a FIQ total score from 0 to < 39 was found to represent a mild effect, ≥ 39 to < 59 a moderate effect, and ≥ 59 to 100 a severe effect.

Conclusion. The analysis indicates that a 14% change in the FIQ total score is clinically relevant, and results of these analyses should enhance the clinical utility of the FIQ in research and practice.

(First Release April 15 2009; J Rheumatol 2009;36:1304–11; doi:10.3899/jrheum.081090)

Key Indexing Terms:
FIBROMYALGIA  FIBROMYALGIA IMPACT QUESTIONNAIRE  CLINICAL TRIAL
MINIMAL CLINICALLY IMPORTANT DIFFERENCE  SEVERITY  PAIN

Fibromyalgia (FM) is a disorder characterized by widespread pain for at least 3 months that affects 3 or 4 quadrants of the body, including axial distribution. Although widespread pain and multiple tender points are the hallmarks of FM, other symptoms such as sleep disturbance, fatigue, stiffness, cognitive problems, headache, anxiety, irritable bowel, and bladder problems are also very common in patients with this disorder. Patients with FM also report that impairments in functional capacity and quality of life are major ongoing concerns.

The Fibromyalgia Impact Questionnaire (FIQ) is a validated, disease-specific composite measure that was developed to determine the spectrum of problems related to FM and responses to therapeutic intervention. It was modified in 1997 and 2002 to reflect experience with using the instrument and to clarify the scoring system. The FIQ is composed of 10 questions (Appendix) and is based on recall in the past week. The first question contains 11 items related to the ability to perform large-muscle tasks with each question rated on a 4-point Likert-type scale. Questions 2 and 3 ask patients to mark the number of days they felt well and the number of days they were unable to work (including household) because of FM symptoms. Questions 4 through 10 are horizontal linear visual analog scales (VAS) marked in 10 increments on which the patient rates work difficulty, pain, morning tiredness, stiffness, anxiety, and depression. The scoring of the FIQ total (0–100) is such that a higher score indicates a greater impact of FM on the person. An average patient with FM has a total score of 50, and severely impaired patients have a total score of 70 or more. Each of the 10 questions is scored 0–10, with a higher score also representing greater impairment.

The FIQ has been translated and validated in several languages and has been used extensively in FM studies, being cited in more than 250 publications. It is a sensitive measurement of change in symptomatology in FM and has been shown to be a more responsive measure of patient-perceived improvement than changes in pain intensity, tender point count, and total tender point pain. While there are studies that...
have reported statistically significant changes of FIQ scores associated with treatment8-12, the clinical relevance of such changes has not been systematically evaluated. We examined within-group change between pairs of adjacent categories on a meaningful external measure using an anchor-based approach13,14 to determine what might be considered a clinically important change in the FIQ total score, using pooled data from patients with FM treated in 3 clinical trials with pregabalin, a US Food and Drug Administration-approved treatment for FM. Given that stiffness is a common complaint among patients with FM1,2,15, we also evaluated clinically important change in the FIQ stiffness item.

MATERIALS AND METHODS

Studies included in the analysis. Data from 3 double-blind, randomized, placebo-controlled, parallel-group clinical trials in patients with FM were included in the analysis. All 3 studies evaluated fixed doses of pregabalin 300, 450, and 600 mg/day using twice-daily dosing (Study 116, Study 212, Study 317). Study 1 evaluated pregabalin treatment effects over 13 weeks16 and the other 2 studies over 14 weeks12,17. The 3 studies were similar in design except for the two 14-week studies, which were preceded by a 1-week, single-blind, placebo run-in period, after which patients with a ≥30% reduction in the pain visual analog scale (VAS) were excluded12,17. Further details on the designs of the studies have been described12,16,17. The studies were conducted in accord with the Declaration of Helsinki, and local regulations and protocols were approved by institutional review boards or independent ethics committees.

Patients. All patients gave written informed consent. Patients included in these 3 studies were women or men aged ≥18 years who met the American College of Rheumatology (ACR) criteria for FM1. At both screening and randomization patients had a score of ≥ 40 mm on a 100-mm pain VAS and had an average score of ≥ 4 on the daily pain diary (11-point pain rating scale, 0 = no pain to 10 = worst possible pain) based on at least 4 entries in the week before randomization. Patients with any active inflammatory disorders or painful conditions that might confound the assessment of FM-related pain were excluded, as were those with unstable medical disorders or a creatinine clearance of ≤ 60 ml/min. Patients who, in the opinion of the investigator, had clinically significant psychiatric conditions, including severe depression, were also excluded. Patients were required to discontinue medications taken for pain and sleep disorders as well as any other psychotropics at least a week before being randomized to placebo or pregabalin monotherapy. Acetaminophen, up to 4 g/day, was the only rescue analgesia permitted during the studies and patients taking low-dose aspirin for cardiac prophylaxis were allowed to continue treatment.

Assessments included in the analysis. The FIQ was completed at baseline and endpoint in all 3 studies and was also completed at Weeks 5 and 9 in Study 1. A global assessment tool, the Patient Global Impression of Change (PGIC), was completed at endpoint (wk 13/14 or at early discontinuation in each study) and at Week 5 in Study 1. Patients were asked to rate their change of overall status on a 7-point scale ranging from 1 = “very much improved” to 7 = “very much worse”18. The PGIC provides an overall assessment of the patients’ own perception of improvement or worsening in pain and other symptoms in conjunction with the influence of treatment side effects, and is used to evaluate the clinical significance of the treatment effect19,20. Using a pain diary, patients were asked to rate their pain in the past 24 h on an 11-point numerical scale. The pain diary was completed each morning upon awakening. The average of the last 7 daily entries was used to determine the baseline, weekly, and endpoint pain score for each patient.

Data analysis. For each study, the estimated mean percentage change in both the FIQ total score and the FIQ stiffness item score for each of the 7 PGIC categories was determined. To derive the minimal clinically important difference (MCID) a repeated measures model was used to estimate the relationship between the percentage change in the FIQ total and stiffness scores and the PGIC using SAS Proc Mixed21. The model provided average estimates of changes in the FIQ scores that corresponded to 1-category differences on the PGIC, i.e., the difference between any 2 adjacent categories. Calculating the MCID automatically accounts for each patient’s baseline score. Using a global rating scale, such as the PGIC, as an external criterion to evaluate the clinical relevance of change is credible22 and has been employed in FM23 and in other conditions, including painful conditions23-26. The percentages of patients in each of the treatment groups who were considered “responders” according to the MCID yielded by the analysis were also calculated, and the statistical significance of the differences between each of the pregabalin dose groups and the placebo group were estimated using bootstrapped simulations27. The 95% confidence intervals (95% CI) were obtained by nonparametric bootstrapping with 50,000 replications.

To estimate severity cutoffs for the FIQ, we first analyzed correlation patterns between the FIQ and a scale with established severity categories, the VAS pain scale. The correlation between the FIQ total score and the pain diary score was calculated for each study and pooled across studies by applying Fisher’s Z transformation to the Pearson product-moment correlation coefficient and its inferences (95% CI and p values)23, using SAS Proc Mixed. Both baseline (pretreatment) and subsequent values (on-treatment) were included. The correlation between the FIQ pain item and the average pain score based on the daily pain diary was also evaluated across studies at each FIQ assessment.

As strong correlations were found between the FIQ total score and the pain diary, it was reasonable to determine FIQ severity categories scores using pain severity as an anchor. An analysis of data from patients with diabetic peripheral neuropathy employed the original method described by Serlin, et al20 to determine optimal cutpoints for pain severity categories on the 11-point numerical rating scale20. The optimal cutpoints were found to be 0–3 = mild, 4–6 = moderate, and 7–10 = severe29. To create an uninterrupted FIQ severity scale we used values of 3.5 and 6.5 on the pain scale as likely boundaries between pain severity categories (as pain averaged over time is a continuous variable, not just an integer from 0 to 10). A repeated measures model was used to estimate the relationship between the FIQ total and pain scores using Proc Mixed in SAS21. The association between treatment and FIQ severity was assessed with the Pearson chi-squared statistic31.

RESULTS

Patients and efficacy assessments. The baseline characteristics were similar across the 3 studies and across treatment groups within each study, as described12,16,17. In total, 2,228 patients were included in the pooled analysis, of whom 93% were women. The mean age was 49 years, patients had been diagnosed with FM for an average of 9.3 years and had a mean of 17.1 tender points. The mean [standard deviation (SD)] baseline pain score was 6.8 (1.3), the mean FIQ total score was 61.7 (14.7), and the mean FIQ stiffness item score was 7.6 (1.9). In total, PGIC assessments were completed by 2,026 patients at endpoint. Among these patients they rated themselves at endpoint as follows: very much worse 3%, much worse 7%, minimally worse 7%; no change 17%, minimally improved 28%, much improved 27%, and very much improved 11%.

Clinically important percentage change. The percentage change in the FIQ total score in each PGIC category was similar in each of the 3 studies (Figure 1). The MCID for the FIQ total score was also similar across each of the 3 studies (Table 1). Overall, pooled across the 3 studies, the estimated MCID...
Figure 1. Mean percentage change in Fibromyalgia Impact Questionnaire (FIQ) total (A) and stiffness (B) item scores for each Patient Global Impression of Change (PGIC) category in each study and for all studies pooled. Patients were asked to rate the PGIC at the end of the study as follows: Since the start of the study, my overall status is: 1 = very much improved; 2 = much improved; 3 = minimally improved; 4 = no change; 5 = minimally worse; 6 = much worse; 7 = very much worse.
in the FIQ total score (0–100) was 14.1% (95% CI 13.3%; 14.9%). For the FIQ stiffness item (0–10), Study 3 gave higher mean percentage change for each PGIC category than Studies 1 and 2 did, but the difference in mean percentage change between each pair of PGIC categories was comparable for the 3 studies. Thus, the clinically important percentage changes, expressed as the difference in mean percentage change for a 1-category change on the PGIC, were similar on the FIQ stiffness item as well. Overall, pooled across the 3 studies, the estimated MCID in the FIQ stiffness item was 13.2% (95% CI 11.9%; 14.4%).

The percentages of patients who achieved a 14% reduction (improvement) in the FIQ total score between baseline and endpoint were as follows: 46% in the placebo group, 48% pregabalin 300 mg/day, 54% pregabalin 450 mg/day, and 49% pregabalin 600 mg/day. Bootstrap analysis of the comparison between the pregabalin groups and the placebo group showed that the difference between the pregabalin 450 mg/day group and placebo was 12% (95% CI 1.8%; 22.1%), a significant finding as the 95% CI did not include zero. The differences between the pregabalin 300 mg/day and 600 mg/day groups were 2.8% (95% CI –3.2%; 8.7%) and 3.6% (95% CI –2.3%; 9.4%), respectively, and not statistically significant. The MCID is that it accounts for the varying baseline levels of functional impairment. By calculating the percentage change in the FIQ total score from baseline for each patient and linking this to each patient’s PGIC category we were able to estimate the relationship between them. Using such an anchor-based interpretation of change in a functional score is a preferred method of evaluating clinical relevance, as it is a readily understood clinical phenomenon, and avoids elaborate statistical procedures. To include a heterogeneous and diverse array of responses across the range of FIQ and PGIC scores to identify an accurate relationship, we combined responses across both treatment groups (pregabalin and placebo). The results from each of the 3 studies individually were generally similar and in close agreement. We found that a 14% change in the FIQ total score could be considered a MCID. Some researchers may prefer to use the absolute values instead of percentages, or in addition to percentages, to define a clinically meaningful difference. If we applied the same methodology for the absolute change as we did for the percentage change, then clinically meaningful differences of 8.1 (95% CI 7.6; 8.5) and 0.89 (95% CI 0.82; 0.96) would result for the FIQ total score and FIQ stiffness item, respectively.

We used an anchor-based approach — with PGIC as the

**Table 2.** Fibromyalgia Impact Questionnaire (FIQ) total score severity categorization using pain severity as an anchor (pooled data).

<table>
<thead>
<tr>
<th>Pain Severity (anchor)</th>
<th>FM Impairment Category</th>
<th>FIQ Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3.5</td>
<td>Mild impairment</td>
<td>&lt; 39</td>
</tr>
<tr>
<td>3.5–6.5</td>
<td>Moderate impairment</td>
<td>≥ 39 to &lt; 59</td>
</tr>
<tr>
<td>≥ 6.5</td>
<td>Severe impairment</td>
<td>≥ 59</td>
</tr>
</tbody>
</table>

**DISCUSSION**

We analyzed clinical data from over 2,000 patients with established, ACR-defined FM and clinically relevant pain as an inclusion criterion. The mean baseline total FIQ score of 62 was consistent with these patients’ moderately severe FM symptomatology. Patients were treated and followed for 3 months, enabling us to evaluate a clinically relevant improvement in the FIQ total score, and to obtain insight into severity categorization using the FIQ. The data were derived from 3 placebo-controlled clinical trials of pregabalin at fixed doses of 300, 450, and 600 mg/day. Analyses were undertaken to gauge the MCID of the FIQ total score and the FIQ stiffness item.

The MCID methodology is intended to quantify whether an individual patient within a particular treatment group has a clinically significant response and can be used to assess the mean percentage change within a particular treatment group. One advantage of using the MCID is that it accounts for the varying baseline levels of functional impairment. By calculating the percentage change in the FIQ total score from baseline for each patient and linking this to each patient’s PGIC category we were able to estimate the relationship between them. Using such an anchor-based interpretation of change in a functional score is a preferred method of evaluating clinical relevance, as it is a readily understood clinical phenomenon, and avoids elaborate statistical procedures. To include a heterogeneous and diverse array of responses across the range of FIQ and PGIC scores to identify an accurate relationship, we combined responses across both treatment groups (pregabalin and placebo). The results from each of the 3 studies individually were generally similar and in close agreement. We found that a 14% change in the FIQ total score could be considered a MCID. Some researchers may prefer to use the absolute values instead of percentages, or in addition to percentages, to define a clinically meaningful difference. If we applied the same methodology for the absolute change as we did for the percentage change, then clinically meaningful differences of 8.1 (95% CI 7.6; 8.5) and 0.89 (95% CI 0.82; 0.96) would result for the FIQ total score and FIQ stiffness item, respectively.
anchor and FIQ as the targeted measure — to examine clinically important change. The anchor approach is consistent with that advocated by others for defining a clinically important change\textsuperscript{13,14,23}. The 2 essential characteristics of a good anchor are that it (1) be interpretable and (2) share an appreciable correlation with the target measure. Both of these characteristics were met in our investigation. Different ways of using the anchor approach have been proposed. For example, one way to define a clinically important improvement in pain has been to collapse categories on “much improved or very much improved” on PGIC\textsuperscript{23}. In the analyses presented in this article, we made each of the 7 categories on PGIC stand on its own merit and took the average difference between any pair of adjacent categories on PGIC as an estimate of the clinically important difference on the FIQ (which holds for clinical worsening as well as clinical improvement). In doing so, we avoided the arbitrariness of collapsing certain categories and capitalized on all available responses across the entire range of the PGIC.

One potential limitation is the assumption of a linear relationship between the FIQ and the 7 PGIC categories. In particular, FM does not appear to be a progressive disease and

![Figure 2. Distribution of patients according to baseline (BL) and endpoint (EP) FIQ severity categories.](image)

<table>
<thead>
<tr>
<th>Severity</th>
<th>Placebo n = 558</th>
<th>300 mg Group, n = 551</th>
<th>450 mg Group, n = 554</th>
<th>600 mg Group, n = 564</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>26.88</td>
<td>27.77</td>
<td>34.30</td>
<td>32.62</td>
</tr>
<tr>
<td>Difference from placebo (95% CI)</td>
<td>—</td>
<td>0.89 (−4.57; 6.14)</td>
<td>7.42 (2.17; 12.9)</td>
<td>5.74 (0.44; 11.09)</td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>32.44</td>
<td>36.48</td>
<td>35.02</td>
<td>34.04</td>
</tr>
<tr>
<td>Difference from placebo (95% CI)</td>
<td>—</td>
<td>4.04 (−1.56; 9.66)</td>
<td>2.58 (−3.05; 8.07)</td>
<td>1.6 (−3.83; 7.04)</td>
</tr>
<tr>
<td>Severe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>40.68</td>
<td>35.75</td>
<td>30.69</td>
<td>33.33</td>
</tr>
<tr>
<td>Difference from placebo (95% CI)</td>
<td>—</td>
<td>−4.93 (−10.7; 0.77)</td>
<td>−9.99 (−15.63; −4.47)</td>
<td>−7.35 (−12.96; −1.69)</td>
</tr>
</tbody>
</table>
comes34,35. We found that most patients were rated as severe scores clinically important for other patient-reported outcomes in assessing treatment differences, as a criterion for study to represent severe impairment. The severity bands can be used in quite good agreement with the approximations when the total score < 39 could be considered to represent a mild impairment, scores ≥ 39 to < 59 to represent a moderate effect, and a score ≥ 59 to represent a severe effect. This analysis is in quite good agreement with the approximations when the scale was originally developed, in which scores ≥ 70 were said to represent severe impairment. The severity bands can be used in assessing treatment differences, as a criterion for study inclusion, and even to serve as an anchor to define change scores clinically important for other patient-reported outcomes. We found that most patients were rated as severe on the FIQ at baseline, which was not surprising given that the inclusion criteria required a score of 4 or greater (moderate to severe) on the pain VAS. At endpoint, there was a notable reduction in the percentage of patients rated as severe and an increase in the percentage of patients rated as mild.

The apparent gender bias in the FIQ, which results from it being developed in a predominantly female population and including items that are more likely to be performed by women, does not affect the legitimacy of our findings as most of this sample also consisted of women. Our findings might not hold true in men with FM. As patients with severe depression or other psychiatric disorders were excluded from the studies, our findings may not extend to this subgroup. A further consideration is that these analyses were based on change over a 3-month period, but might not be applicable to longer periods of time; i.e., patients with a 14% improvement after 6 months or a year might not rate themselves as having improved.

These data were derived from FIQ changes in response to a medication; data from a nonpharmacological intervention (e.g., exercise) may yield a different MCID. Patients with severe depression or unstable psychiatric conditions were excluded from the studies; their inclusion may have yielded a different MCID. FM may occur in association with inflammatory disorders such as rheumatoid arthritis; changes in the FIQ have not been evaluated in combined disorders and it is likely that the MCID reported here would not be applicable in such situations. It should also be noted that a MCID may vary in different study populations, type of anchor, time period of assessment, and other considerations; thus, our findings may not be generalizable. Moreover, the estimated MCID defined and derived here refers to a one-category difference on the PGIC, so values of MCID less than those estimated here would result for less than a one-category difference (e.g., a one-half category difference) on the PGIC.

The analyses reported here enhance the clinical utility of the FIQ in practice and assist in the interpretation of findings from clinical trials. In patients with moderate to severe impairment on the FIQ, a change of approximately 14% over 3 months is likely to be clinically important. A similar percentage change in stiffness is also likely to be clinically important.

ACKNOWLEDGMENT
We thank Dr. Bernhardt Zeiher, Pfizer Inc., for his input into the analysis. Editorial support was provided by Dr. Janet Bray and funded by Pfizer Inc.

REFERENCES