A rating scale for fibromyalgia and chronic fatigue syndrome (the FibroFatigue scale)

Olof Zachrisson*, Björn Regland, Marianne Jahreskog, Margareta Kron, Carl G. Gottfries

Psychiatry Section, Institute of Clinical Neurosciences, Göteborg University, Göteborg, Sweden

Received 6 February 2001; accepted 1 November 2001

Abstract

Objective: To construct an observer’s rating scale sensitive to change for measuring severity and treatment outcome in fibromyalgia (FM) and chronic fatigue syndrome (CFS) patients. Methods: A selection of items from the Comprehensive Psychopathological Rating Scale (CPRS) were repeatedly rated and used as outcome measure of a 24-week treatment study. In the study 100 women, fulfilling the criteria for both FM and CFS, received intermittent injections of a staphylococcus toxoid or placebo. Nine CPRS-items with high baseline incidence (cutoff 70%) were extracted and validated against global ratings and the Fibromyalgia Impact Questionnaire (FIQ). The fibromyalgia and chronic fatigue syndrome rating scale (the FibroFatigue scale) was thereafter formed based upon the extracted items and three supplemented ones. The interrater reliability was tested in 27 consecutive patients of both sexes. Results: The FibroFatigue scale is an observer’s rating scale with 12 items measuring pain, muscular tension, fatigue, concentration difficulties, failing memory, irritability, sadness, sleep disturbances, and autonomic disturbances (items derived from the CPRS) and irritable bowel, headache, and subjective experience of infection (new items). There was a statistically significant correlation between the CPRS-extracted items and global ratings as well as with the FIQ. The interrater reliability of the new scale was excellent (correlation coefficient .98), irrespective of the patients’ gender. Conclusion: The FibroFatigue scale seems to be a reliable and valid measuring instrument with capacity to monitor symptom severity and change during treatment of FM/CFS patients. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: Fatigue syndrome; Chronic; Fibromyalgia; Infection; Psychiatric status rating scales; Pain; Treatment outcome

Introduction

Fibromyalgia (FM) and chronic fatigue syndrome (CFS) are both chronic clinical conditions characterised by a variety of nonspecific symptoms, including prominent fatigue, myalgia, and sleep disturbances. Despite different diagnostic criteria, CFS and FM have many demographic and clinical similarities with few differences in the domains of symptoms [1,2]. Both disorders seek their clinical recognition and, so far, their diagnoses are based on criteria. There are no widely accepted pathogenic explanatory models for either illness [2,3]. In spite of extensive research, treatment continues to be of limited success [4].

Treatment studies on FM/CFS have frequently used the Visual Analogue Scale (VAS) to obtain patient perception data. The profile of mood has been estimated using rating scales for depression such as the Hamilton Depression Rating Scale [5] and the Beck Depression Inventory [6]. The Arthritis Impact Measurement Scales [7] and the Fibromyalgia Impact Questionnaire (FIQ) [8] are self-rating instruments which have been used to assess functional limitations and disability. They have been evaluated for reliability and validity in the FM population [9]. Self-rating scales are also provided for the evaluation of fatigue [10–12] and one self-rating scale for screening CFS symptoms has been published [13]. In a previous trial on FM/CFS [14], we found the neurasthenia subscale of the Comprehensive Psychopathological Rating Scale (CPRS) [15,16] useful for evaluation of treatment effects.

Within psychiatry, observer’s rating scales are often considered more valid and reliable than self-rating scales.
In this study, we wanted to construct a new observer’s rating scale aimed at the assessment of core symptoms and treatment outcome in FM and CFS. Due to substantial overlap in symptomatology between FM and CFS, we decided to develop a scale suitable for both conditions. The scale should be based on the CPRS items that we had used within a double-blind treatment study and their psychometric properties be tested against global rating scales [17] and the FIQ.

Patients and ratings

Patients

The present study was based on a group of 100 patients, all women, aged 20–66 years [mean 47.8 years (± 11.1 S.D.)] who were evaluated in a 24-week double-blind clinical trial. In the trial, patients were given repeated subcutaneous injections of a staphylococcus toxoid vaccine or placebo [18]. To be included in the research program, the patients had to fulfill the American College of Rheumatology (ACR) criteria for FM [19] and the 1994 criteria for CFS established by the CDC [20]. The patients were referred to our unit from other units in the hospital but most of them were referred by general practitioners. All patients included were chronically ill and many had been granted a sickness pension. On average, the patients had been suffering from FM symptoms and chronic fatigue for 11.5 years (range 1–40 years). Due to ethical considerations, the patients were allowed to keep to their prescribed drug regimens during the study. The most common medication was analgesics, which were used by 79% of the patients. Antidepressants (low doses of tricyclics or selective serotonin reuptake inhibitors) were used by 42%, hypnotics by 26%, daytime benzodiazepine by 21%, and medications for gastrointestinal problems by 19%. Seven patients (7%) did not use any concomitant medication.

Scales used to design and validate the FibroFatigue scale

The fibromyalgia and chronic fatigue rating scale (the FibroFatigue scale) was derived from 15 items of the neurasthenia subscale of CPRS [15] (Table 1). The ratings were performed within the study by three registered nurses trained by an experienced CPRS rater. An interview was done at baseline, and new interviews were done once a month during the 24-week trial. At every interview, the same rater, who was blinded to whether the patient was given active treatment or placebo, assessed the patient. The duration of the interview varied but was usually less than 20 min.

The original CPRS [21] comprises 65 items covering a broad range of psychiatric symptoms both of neurotic and psychotic dimensions. The scale was intended for treatment evaluation and to serve as a pool from which items could be selected to construct scales for various syndromes. The CPRS is a semistructured interview-based rating instrument selected to construct scales for various syndromes. The evaluation and to serve as a pool from which items could be selected to construct scales for various syndromes.
Method

Primary item selection for the new scale

Of the 15 CPRS items used both in a previous [14] and in the present study, we wanted to exclude those that did not occur sufficiently often to be regarded as core symptoms and those that received so low scores that changes would be difficult to register (floor effect). The baseline frequencies of scores above zero on the 15 items were therefore calculated and ranked by incidence. An arbitrary cutoff point of 70% occurrence was used to identify the most common items in the total sample of patients.

Convergent construct validity of selected items

Total scores and changes in these scores

The sum of scores on the items with at least 70% occurrence was used as a preliminary estimate of symptom severity. This estimate was correlated with the CGI-S and with the total scores on the FIQ.

After baseline assessments, the patients were randomised to receive active treatment (staphylococcus toxoid) or placebo. The difference in total scores on the selected (for the new scale) CPRS items between baseline and Week 24 (endpoint of treatment) was used as an estimate of change in symptomatology during the trial. The same estimate of change was calculated for the FIQ. We also divided the patients into responders and nonresponders according to their CGI-C scores. Patients who were rated as (either of) minimally improved, much improved, or very much improved were categorised as responders and those who were rated as (either of) unchanged, minimally worse, much worse, or very much worse were categorised as nonresponders. The scores on the selected CPRS items before and after treatment were then compared between the responders and nonresponders.

The estimate of change expressed in CPRS scores was validated against ratings on the CGI-C and FIQ using Pearson product–moment correlation.

Scores on individual items and changes in these scores

The selected CPRS items were correlated with comparable items of the FIQ at baseline and at Week 24 in the trial. For each item, the difference between the baseline and Week 24 values was used as an estimate of change. The changes in scores on the selected CPRS items were then correlated with the changes in scores on comparable FIQ items.

Estimates of the selected items’ sensitivity to change

To estimate the sensitivity to change of the individual CPRS items, two different values were calculated for the group of patients that had received active treatment. First the mean change in scores (absolute value) on each item after 24 weeks of treatment was calculated and ranked. Then, for each item, the correlation between the change in scores on this item and the total change in scores on the selected items was analysed. Ideally, an item should both yield large changes (which can be reliably rated) and correlate strongly with the general improvement in symptomatology [24].

Construction of the FibroFatigue scale

The FibroFatigue scale comprises 12 observer-rated items. After validity analyses of the CPRS, nine of these items constituted the main part of the new scale. Based on our clinical experience, we allowed reformulation of four of the items and addition of three newly constructed items to cover relevant symptomatology of the syndromes. In the reformulation and construction of items, care was taken to give the scale the same format as the scale from which it was derived.

Interrater reliability testing of the FibroFatigue scale

The interrater reliability of the FibroFatigue scale was assessed by calculating the correlation coefficients between different raters. The reliability testing would also give us some indications of the usefulness of the modified and new items. A low degree of concordance for an individual item could indicate that the item or an anchor point description was poorly formulated. We also wanted to test the new scale on patients of both sexes. High interrater reliability with respect to ratings of males would support the scale’s usefulness also in this type of population. Conjoint interviews with consecutive patients were therefore conducted using the FibroFatigue scale. Fourteen of the patients were women and 13 were men. They all fulfilled the criteria for both FM and CFS. Two trained nurses and one psychiatrist were present at the interviews and alternated in performing them. All ratings were performed without communication between raters.

Statistics

Values are expressed as mean and standard deviations (S.D.) if not otherwise stated. The statistical analyses were conducted using paired t test for paired samples and Student’s t test for calculating group differences. All tests were two-tailed. A P value < .05 was considered statistically
significant. Correlations were tested with Pearson product–moment correlation. All statistical comparisons of central tendencies were also checked with nonparametric methods (Wilcoxon’s signed rank test, Mann–Whitney’s U test, and Spearman’s rank correlation). Both types of method yielded the same significance levels. We have therefore chosen only to present the outcomes obtained with parametric methods. The interrater reliability was tested using ANOVA with patients and raters as independent factors as described by Fleiss [28].

Results

Primary item selection for the FibroFatigue scale

One hundred patients were evaluated at baseline. For 90 of them, CPRS scores, as well as scores on the FIQ and the global scales, were available from ratings before and after 24 weeks of treatment. Six of the 15 CPRS items were found to have a baseline incidence below 70%. The mean scores on these items were also low (Table 1). For these reasons, they were excluded from further analysis. Thus, nine items remained for evaluation. These were: “aches and pain,” “fatiguability,” “reduced sleep,” “muscular tension,” “concentration difficulties,” “failing memory,” “autonomic disturbances,” “hostile feelings,” and “sadness.” This 9-item scale was called the CPRS-9.

Convergent construct validity of selected items

Total scores and changes in these scores

At baseline, the correlation between the CPRS-9 total scores (our preliminary estimate of symptom severity) and illness severity measured by the CGI-S was significant ($r = .43, P < .001$). There was also a significant correlation between the CPRS-9 total scores and the FIQ total scores ($r = .48, P < .001$). At the end of treatment, the correlations were slightly higher. The correlation coefficient between the CPRS-9 and CGI-S total scores at Week 24 was .56 ($P < .001$), and the correlation coefficient between the CPRS-9 and FIQ total scores at the same time-point was .54 ($P < .001$).

During the course of the double-blind trial, global ratings on the CGI-C suggested that many patients improved with the treatment. Table 2 illustrates that this was also reflected in the CPRS-9 ratings. In the responder group, as defined by the CGI-C, there was a significant reduction, 38%, in the CPRS-9 total score after 24 weeks of treatment ($t = 10.6, df = 39, P < .001$). In the group of nonresponders, there was no significant change in CPRS-9 total scores during the treatment period ($t = 0.06, df = 49, P = .95$).

The change in total scores on the CPRS-9 showed a significant correlation with the global judgements on the CGI-C ($r = .70$) and FIQ total scores ($r = .67$).

Scores on individual items and changes in these scores

When comparing individual items of the CPRS-9 and FIQ, the observer-rated items of “aches and pain,” “fatiguability,” “sadness,” and “reduced sleep” all showed significant correlations (.67, .67, .57, and .56) with their respective self-reported visual analogue FIQ scales after 24 weeks of treatment. The correlations were also statistically significant at baseline but within a lower range.

Comparing changes in scores, the CPRS items of “aches and pain,” “fatiguability,” “sadness,” and “reduced sleep” showed significant correlations (.61, .62, .51, and .42) with the FIQ analogues.

Estimates of the sensitivity to change of the CPRS-9

For each of the nine CPRS items, the mean change in scores in the group receiving active treatment was calculated and ranked (Table 3). The item of “fatiguability” was found to be the most reduced with a mean change in scores of 1.56 ($± 1.73$ S.D.), followed by “reduced sleep” with a mean change in scores of 1.36 ($± 1.52$ S.D.). The lowest change

### Table 2

<table>
<thead>
<tr>
<th>Classification according to CGI-C</th>
<th>Total score before treatment (mean ± S.D.)</th>
<th>Total score after treatment (mean ± S.D.)</th>
<th>Level of significance, paired t test ($P$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients who improved $n = 40$</td>
<td>28.9 ± 5.3</td>
<td>17.9 ± 7.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Patients who did not improve $n = 50$</td>
<td>29.0 ± 5.3</td>
<td>28.8 ± 7.2</td>
<td>n.s.</td>
</tr>
<tr>
<td>$P$</td>
<td>n.s.</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

The patients were women fulfilling the diagnostic criteria for FM and CFS.
in scores was found in “muscular tension” with a mean of 0.64 (± 1.42 S.D.). The changes in scores on the individual items all showed significant correlations with the change in total scores. The strongest correlation with the change in total scores was found in “aches and pain” and “fatiguability” ($r = .85$ and $.86$, respectively; $P < .001$), the weakest between the change in total scores and “hostile feelings” ($r = .45$; $P < .05$). The latter is an item that was modified in the final version of the scale.

**Modifications of CPRS-9 items included in the new scale**

The FibroFatigue scale comprises 12 observer rated items (see Appendix A) and is intended to reflect symptom severity and be sensitive to change during treatment. The scale measures pain, muscular tension, fatigue, concentration difficulties, failing memory, irritability, sadness, sleep disturbances, and autonomic disturbances (nine items derived from the CPRS) and irritable bowel, headache, and subjective experience of infection (new items). In the FibroFatigue scale the CPRS item “fatiguability” was adjusted to include general fatigue and lack of energy and the name of the item was changed to “fatigue.” “Hostile feelings” was renamed “irritability” and its definition and scale step descriptions were slightly modified. The item of sleep was adjusted to include both insomnia and hypersomnia and was renamed “sleep disturbances.” The definition and scale text in “autonomic disturbances” were modified because we decided to rate irritable bowel symptoms separately.

**Interrater reliability testing of the FibroFatigue scale**

The interrater reliability was tested using ANOVA with patients and raters as independent factors. The correlation coefficient was .98 for the total scale, with a range of .86–.97 for individual items, indicating satisfactory concordance (Table 4). The interrater reliability was high irrespective of whether men or women were rated (correlation coefficient of .99 and .97, respectively, for the total scale).

**Discussion**

To our knowledge, there is no observer’s rating scale available that is specifically designed for FM or CFS. Despite different diagnostic criteria, these two conditions have many demographic and clinical similarities [1,2]. They are both characterised by a variety of nonspecific symptoms, including prominent fatigue, myalgia, and sleep disturbances. Up to 70% of patients with CFS-like illnesses may have concurrent FM [29] and it may be of little importance to differentiate them [30]. On the basis of a literature review, Wessely et al. [31] found a substantial overlap, even within the entire group of functional somatic syndromes, including FM, CFS, irritable bowel syndrome, and multiple chemical sensitivity, among others. In this study, we chose to construct and evaluate a new rating scale that would apply to both FM and CFS. The development of the scale was made possible due to ratings on CPRS items in a double-blind trial where these items could be tested against global measures and the FIQ [18]. The FibroFatigue scale is not intended for diagnosis. Nor is the scale intended to be a screening instrument. Instead, our intention was to develop an instrument that could assess symptom severity and change over time and that was suitable for use in clinical trials as well as in the clinic.

In the trial, 15 CPRS items were selected by experienced clinicians as having face validity with the disorders. Nevertheless, six of these items were excluded in the final scale because of low incidence at baseline (low content validity). The anxiety items of “inner tension,” “worrying over trifles,” “hypochondriasis,” and “phobias” all showed an incidence below 60%. This was somewhat surprising, since, in the ACR criteria [19], anxiety is classified as a moderately common symptom. A lifetime history of depression has been reported in 50% to 70% of patients with FM [32]. Current major depression is less common, however, and most patients with FM are not depressed. A high rate of depressive comorbidity has been reported in CFS [2]. The FibroFatigue scale has kept the item of “sadness,” but the items of “pessimistic thoughts” and “suicidal thoughts” were excluded because of low incidence. Ethical considerations made drug withdrawal impossible and therefore treatment with antidepressants may be a confounding factor in the rating of mood.

The content validity of the items of aches and pain, fatiguability, reduced sleep, concentration difficulties, failing memory, and autonomic disturbances is in line with the diagnostic criteria [19,20]. Headache is included in the
diagnostic criteria for CFS and is frequently seen in FM [32]. The comorbidity with irritable bowel syndrome is frequent in both conditions [33,34]. Symptoms of infection are included in the criteria for CFS, and certain infections may trigger FM [35]. Furthermore, according to our clinical experience, muscular tension and irritability are frequently occurring symptoms in these disorders.

The items selected for testing were effective in discriminating responders from nonresponders, and there were significant correlations between the change in scores (improvement) on the CPRS-9, the global change, and the FIQ scores, indicating proper validity with regard to measurement of change. Based on the test results, the nine CPRS items were included in the FibroFatigue scale (four of them after some reformulation), and three newly constructed items were added.

In the population studied, the FibroFatigue scale was found to be sensitive to change during a treatment period. Tender-point count has been tried in FM but, in reported studies, tender-point scores did not correlate significantly with symptoms or treatment response [36,37]. It will be interesting to see how the scale functions in pure diagnostic groups, with either FM or CFS patients. A possible criticism against the scale would be that it combines two conditions. We claim, however, that the scale includes items that are represented in both conditions. An important question is gender differences in the symptomatology. According to the interrater reliability testing, the correlations between different raters were just as high in men as in women. Thus, the scale applies to patients of both sexes. Further potential shortcomings of the study would be the addition of three items to the scale which only were include in the test of reliability. Moreover, as with every observer assessment instruments, even if the reliability is high, there is always a risk for bias if the observer were to be unblinded to treatment.

In conclusion, there is a lack of specific rating instruments for use in FM and CFS. This study presents a new 12-item observer-rating instrument for assessment of symptom severity and change during treatment in these conditions. Items were selected from previously used instruments and the new scale was evaluated for convergent construct validity, interrater reliability, and sensitivity to change. The tested validity and reliability is satisfactory but further studies are needed to investigate the usefulness of the scale.

Acknowledgments

Thanks to Mrs Ingrid Larsson for language revision and linguistic advice.

Appendix A.

The rating applies to reported symptoms and should be made by a medically educated observer. The rating should be based on a clinical interview moving from broadly phrased questions about symptoms to more detailed questions that allow precise rating of severity. The rater must decide whether the rating lies on the defined scale steps (0, 2, 4, 6) or between them (1, 3, 5). The scale may be used for any time interval between ratings, be it weekly or otherwise, but the interval must be recorded.

Item list

1. Aches and pain
2. Muscular tension
3. Fatigue
4. Concentration difficulties
5. Failing memory
6. Irritability
7. Sadness
8. Sleep disturbances
9. Autonomic disturbances
10. Irritable bowel
11. Headache
12. Subjective experience of infection

A.1. Aches and pain

Representing reports of bodily discomfort, aches and pain. Rate according to intensity, frequency, duration, and requests for relief. Disregard any statement about the cause being organic.

0 Absent or transient aches
1 Occasional definite aches and pain
2 Prolonged and inconvenient aches and pain; requests for effective analgesics
3 Severely interfering or crippling pains

A.2. Muscular tension

Representing the description of increased tension in the muscles and difficulty in relaxing physically.

0 No increase in muscular tension
1 Some occasional increase in muscular tension, more evident in demanding situations
2 Considerable difficulty in finding a comfortable position when sitting or lying; disturbing muscular tension
3 Painful muscular tension; completely incapable of relaxing physically
4 Painful muscular tension; completely incapable of relaxing physically
A.3. Fatigue

Representing the experience of debilitating fatigue and lack of energy and the experience of tiring more easily than usual.

0 Ordinary staying power; not easily fatigued
1 Tires easily but does not have to take a break more often than usual
2 Considerable fatigue and lack of energy; easily wearied; frequently forced to pause or rest
3 Exhaustion interrupts almost all activities or even makes them impossible

A.4. Concentration difficulties

Representing difficulties in collecting one’s thoughts mounting to incapacitating lack of concentration.

Rate according to intensity, frequency, and degree of incapacity.

0 No difficulties in concentrating
1 Occasional difficulties in collecting thoughts
2 Difficulties in concentrating and sustaining thought which interfere with reading or conversation
3 Incapacitating lack of concentration

A.5. Failing memory

Representing subjective disturbances of recall compared with previous ability.

0 Memory as usual
1 Occasional increased lapses of memory
2 Reports of socially inconvenient or disturbing loss of memory
3 Complaints of complete inability to remember

A.6. Irritability

Representing the subjective experience of irritable mood (dysphoria), anger, and having a short fuse, regardless of whether the feelings are acted out or not.

Rate according to intensity, frequency, and the amount of provocation tolerated.

0 Not easily irritated
1 Easily irritated or angered; reports irritability, which is easily dissipated
2 Pervasive feelings of irritability or anger; outbursts may occur
3 Persistent irritability or anger which is difficult or impossible to control

A.7. Sadness

Representing subjectively experienced mood, regardless of whether it is reflected in appearance or not; includes depressed mood, low spirits, despondency, and the feeling of being beyond help and without hope.

Rate according to intensity, duration, and the extent to which the mood is influenced by events.

0 Occasional sadness may occur in the circumstances
1 Predominant feelings of sadness, but brighter moments occur
2 Pervasive feelings of sadness or gloominess; the mood is hardly influenced by external circumstances
3 Continuous experience of misery or extreme despondency

A.8. Sleep disturbances

Representing a subjective experience of disturbed sleep compared to the subject’s own normal pattern when well.

0 Sleeps as usual
1 Slight difficulty dropping off to sleep, reduced duration of sleep, light or fitful sleep, or sleeps deeper or longer than usual
A.9. Autonomic disturbances

Representing descriptions of palpitations, breathing difficulties, dizziness, increased sweating, cold hands and feet, dry mouth, and frequent micturition.

Rate according to intensity and frequency and duration of one or many symptoms.

0 No autonomic disturbances
1 Occasional autonomic symptoms which occur under emotional stress
2 Frequent or intense autonomic disturbances (two or more of above-mentioned symptoms) which are experienced as discomforting or socially inconvenient
3 Very frequent autonomic disturbances, which interrupt other activities or are incapacitating

A.10. Irritable bowel

Representing a subjective experience of abdominal discomfort or pain along with descriptions of altered stool frequency or diarrhoea/obstipation, bloating or feeling of distension.

Rate according to intensity, frequency, and degree of inconvenience produced.

0 No irritable bowel
1 Occasional irritable bowel symptoms which may occur under emotional stress
2 Frequent or intense irritable bowel, which is experienced as discomforting or socially inconvenient
3 Very frequent irritable bowel, which interrupts other activities or are incapacitating

A.11. Headache

Representing reports of discomfort, aches, and pain at the head.

Rate according to intensity, frequency, duration, and requests for relief. Disregard any statement about the cause being organic.

0 Absent or transient headache
1 Occasional definite headache
2 Prolonged and inconvenient headache; requests for effective analgesics
3 Severe interfering or crippling headache

A.12. Subjective experience of infection

Representing descriptions of palpitations, breathing difficulties, dizziness, increased sweating, cold hands and feet, dry mouth, and frequent micturition.

Rate according to intensity, frequency, and duration and also requests for treatment.

0 No symptoms of infection
1 Occasional definite symptoms of infection
2 Frequent or intense symptoms of infection; requests for treatment
3 Severe interfering or crippling symptoms of infection

References

[10] Krupp LB, LaRocca NG, Mui̇n-Nash J, Steinberg AD. The fatigue


