Prognosis in chronic fatigue syndrome: a prospective study on the natural course

J H M M Vercoulen, C M A Swanink, J F M Fennis, J M D Galama, J W M van der Meer, G Bleijenberg

Abstract

Objective—To determine spontaneous improvement after a follow up interval of 18 months in patients with chronic fatigue syndrome and to identify factors that predict improvement.

Methods—A longitudinal study was used. Of 298 initially assessed self-referred patients fulfilling criteria for chronic fatigue syndrome, 246 patients completed self-report questionnaires at follow up (response rate 83%). A multidimensional assessment method was used, measuring behavioural, emotional, cognitive, and social functioning. Comparison data from 53 healthy subjects matched for age, sex, and educational level were available.

Results—Three per cent of patients reported complete recovery and 17% reported improvement. At follow up, there were considerable problems at work and consumption of medication was high. Subjective improvement was confirmed by dimensional change: at follow up recovered patients had similar scores to healthy subjects and improved patients showed significant improvement on four out of seven outcome measures and had higher scores than healthy subjects in all dimensions. Sociodemographic variables or treatment by specialists and alternative practitioners did not predict improvement. Predictors of improvement were: subjective sense of control over symptoms, less fatigue, shorter duration of complaints, and a relative absence of physical attributions.

Conclusion—The improvement rate in patients with a relatively long duration of complaints is small. Psychological factors are related to improvement, especially cognitive factors.

(J Neurol Neurosurg Psychiatry 1996;60:489–494)

Keywords: prognosis; chronic fatigue syndrome

Although the number of studies on chronic fatigue syndrome have increased dramatically in the past years, little research has been done on the course of the complaint and on factors associated with its persistence. There are four follow up studies of patients with chronic fatigue syndrome. All patients in the study of Wilson et al4 and most patients in that of Bonner et al2 had been in specific treatment programmes for chronic fatigue syndrome and therefore these studies did not investigate the natural course. The study of Sharpe et al3 assessed variables associated with functional impairment at follow up only. In a study by Clark et al1 patients were referred to a chronic fatigue clinic. It is not clear whether this was a study on the natural course of the disease or that these patients had been treated during the follow up interval. Also, this study did not investigate cognitive factors as possible predictors of improvement.

Detailed prospective studies on the natural course in chronic fatigue syndrome are needed to provide information, not only on prognosis, but also on factors associated with chronicity. Identification of these factors is important to the development of therapeutic interventions.

In a previous study, we developed and tested a multidimensional assessment method for chronic fatigue syndrome, assessing behavioural, emotional, cognitive, and social functioning. The dimensions proved to be relatively independent: each dimension provided a unique contribution to the description of the patient. This means that comprehensive assessment of chronic fatigue syndrome entails measurement on all dimensions simultaneously. Evaluating prognosis, therefore, should also be based on multidimensional assessment.

In the present study, we repeated the multidimensional assessment in the same group of patients 18 months later. The main questions were: Has improvement occurred 18 months later and what factors predict improvement?

Subjects and methods

SUBJECTS

At initial assessment, 357 self-referred patients completed postal questionnaires assessing the dimensions of chronic fatigue syndrome. Of those patients, 298 fulfilled criteria for the disease6 and were included in the study. All patients were seen by at least one specialist. A detailed description of this sample and of initial assessments has been reported previously.3

These patients received the same postal questionnaires 18 months later. Non-responders were sent reminders first, and in a second stage were contacted by telephone. Two hundred and forty six patients finally completed and returned the questionnaires (response rate 83%). Nearly all non-responders had moved during the follow up interval without reporting
their new address, one patient had died, and the remaining patients were not interested in participating anymore. Mean age in the follow up sample was 39 (95% confidence interval (95% CI) 37-41) years; there were 59 men and 187 women. Mean duration of complaints at initial assessment was 8.4 years (95% CI 7.3-9.6; median 4.5). Comparison data of 53 healthy subjects matched for age, sex, and education were available. These subjects, selected through a regional newspaper advertisement, were included to provide a standard to evaluate whether patients who reported improvement or recovery actually had dimension scores similar to healthy subjects.

INSTRUMENTS
The following dimensions were measured. (Details about the construction of dimension scores and the instruments used are described elsewhere.9)

Subjective experience of the personal situation
This dimension reflects the subjective experience of fatigue and disability and includes: general questions on problems in housekeeping activities and work (five point scale) and satisfaction in housekeeping activities and work (five point scale), and the subjective fatigue subscale and the physical activity subscale of the fatigue questionnaire checklist individual strength (CIS).

Psychological wellbeing
The Beck depression inventory (BDI)7 is a standardised self report questionnaire to measure depression. The following subscales were used from the symptom checklist (SCL-90),8 which is a 90 item indicator of psychopathology: anxiety, agoraphobia, depression, somatisation, cognitive difficulties, interpersonal sensitivity.

Functional impairment
The following subscales of the sickness impact profile (SIP) were included10: activities at home, mobility, and ambulation.

Sleep disturbances
These were measured by a sleep disturbances question, which was coded "yes" or "no" and the subscale on sleeping problems of the SCL-90.

Avoidance behaviour
Four questions were asked on avoidance of physical activity as a way of coping with complaints (for example, "avoiding complaints by non-activity"). At initial assessment patients had to indicate whether they did or did not engage in these behaviours, and a total score was derived. In the present study these items were scored on a four point scale (seldom; now and then; often; very often). These items were summed into a total score.

Neuropsychological functioning
This dimension was measured by the self report questionnaires CIS subscale concentration and the SIP subscale concentration.

Social interactions
General questions were asked concerning: (a) satisfaction in social life (five point scale); and (b) problems in social relations (four point scale). Also, the subscale social interactions of the SIP is included.

Sense of control over symptoms (self efficacy expectations)
This dimension contains a specific 3 point scale question measuring the subjective sense of control over symptoms ("Do you think you can influence your complaints?") and the subscale internal attributions of the multidimensional health locus of control questionnaire (health is determined by own behaviour/capacities).11 12

Causal attributions
At initial assessment, patients were asked to write down their opinion regarding the cause of their complaints. The responses were coded into a variable with categories "physical" and "non-physical". In the present study eight possible causes were presented. Patients could indicate on a five point scale if they agreed or disagreed with each cause. Factor analyses yielded two subscales: "physical" and "non-physical". Per subscale item scores were summed and divided by the number of items in the scale.

The following questions were also asked with respect to the follow up interval of 18 months: the patient's opinion of (self reported) change (completely recovered; improved; unchanged; complaints got worse); changes in marital status and occupational situation; visits and treatment by specialists and alternative practitioners for chronic fatigue syndrome; use of medication for chronic fatigue syndrome.

Severity of fatigue and the non-cognitive dimensions (avoidance of physical activity—being measured differently at initial and follow up assessment—was excluded) were considered as the seven outcome measures of clinical status and disability.

STATISTICS
Patients who had missing values in more than one dimension were not included. Missing values were replaced for the variable "causal attributions" at initial assessment, because this variable had missing values in more than 5% of patients. The linear trend at point method was used. The existing series was regressed on an index variable scaled 1 to n. Missing values were replaced with their predicted value. Evaluating relations between variables was performed by $\chi^2$ test in the case of dichotomous or nominal variables. Spearman rank correlation coefficient and Mann-Whitney U test were used for other variables. To evaluate whether there were significant differences between recovered, improved, or non-improved patients in the degree of improvement on the outcome measures of clinical status and disability Mann-Whitney U tests were performed on the difference between scores at initial and follow up assessment.
Bonferroni correction was used for multiple comparisons to control type I error. Prediction of outcome measures was done by stepwise multiple regression analyses. To detect the causal direction of variables at initial assessment that predicted severity of fatigue at follow up, cross lagged panel analyses were performed. The direction of causation between two variables $x$ and $y$ can be inferred by comparing the correlation between variable $x$ on initial assessment and variable $y$ on follow up assessment ($r_{xy Y}$) with the correlation between variable $y$ on initial assessment and variable $x$ on follow up assessment ($r_{yx X}$). The principal direction of causation is said to go from $x$ to $y$ if correlation $r_{xy Y}$ is significantly greater than correlation $r_{yx X}$.

**Results**

**CHANGES IN PERSONAL SITUATION DURING FOLLOW UP INTERVAL**

There were no major changes in marital status (unchanged in 93% of patients). Occupational situation remained unchanged in 197 (80%) patients. Before the onset of complaints, 141 (57%) patients worked. At initial assessment 30 (12%) patients were unemployed, 69 (28%) worked, and 105 (43%) were on sick leave or received disablement insurance benefits. The remaining 42 (17%) patients were housewives, were retired, or went to school. At follow up assessment, 29 (12%) patients were unemployed, 71 (29%) worked, and 103 (42%) were on sick leave or received disablement insurance benefits. The remaining 43 (17%) patients were housewives, were retired, or went to school.

**SELF REPORTED CHANGE**

Eight (3%) patients indicated that they had completely recovered, 42 (17%) that they were feeling much better, 147 (60%) that complaints had remained the same, and 49 (20%) that complaints had become worse.

**Table 1** shows data on the seven outcome measures and depression at initial assessment (A1) and follow up (FU) for eight recovered patients, 42 improved patients, 197 non-improved patients, and 53 healthy subjects. High scores indicate abnormal functioning.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Recovered</th>
<th>Improved</th>
<th>Non-improved</th>
<th>Healthy subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>5.4 (1.8)</td>
<td>5.6 (1.1)</td>
<td>6.2 (0.9)</td>
<td>2.2 (1.3)</td>
</tr>
<tr>
<td>Psychological wellbeing</td>
<td>1.8 (1.0)**</td>
<td>4.5 (1.3)**</td>
<td>6.2 (0.8)</td>
<td>2.2 (1.3)</td>
</tr>
<tr>
<td>A1</td>
<td>69 (37.9)</td>
<td>83.4 (44.2)</td>
<td>86.7 (45.3)</td>
<td>3.8 (15.3)</td>
</tr>
<tr>
<td>FU</td>
<td>24.4 (21.7)**</td>
<td>69.8 (32.2)**</td>
<td>77.4 (44.4)*</td>
<td>21.9 (19.9)</td>
</tr>
<tr>
<td>Functional impairment</td>
<td>7.3 (8.5)</td>
<td>7.9 (5.6)</td>
<td>11.4 (7.9)</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>6.6 (1.0)*</td>
<td>6.3 (5.5)</td>
<td>11.4 (7.7)</td>
<td>0.2 (0.7)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>23.6 (22.2)</td>
<td>6.3 (24.8)</td>
<td>4.0 (21.1)</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>10.9 (12.7)</td>
<td>30.3 (22.3)</td>
<td>42.2 (23.9)</td>
<td>21.5 (15.2)</td>
</tr>
<tr>
<td>FU</td>
<td>5.8 (4.2)</td>
<td>5.2 (4.1)</td>
<td>5.9 (3.7)</td>
<td>2.9 (2.8)</td>
</tr>
<tr>
<td>Sleep disturbances</td>
<td>34.6 (3.6)</td>
<td>4.4 (3.6)</td>
<td>5.9 (3.5)</td>
<td>2.9 (2.8)</td>
</tr>
<tr>
<td>Concentration</td>
<td>18.2 (11.9)</td>
<td>35.7 (18.0)</td>
<td>34.1 (16.3)</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>2.9 (2.5)**</td>
<td>26.0 (17.1)*</td>
<td>34.1 (16.1)</td>
<td>5.6 (5.0)</td>
</tr>
<tr>
<td>Subjective experience</td>
<td>35.6 (17.2)</td>
<td>40.8 (10.8)</td>
<td>44.8 (10.6)</td>
<td>11.8 (10.4)</td>
</tr>
<tr>
<td>FU</td>
<td>4.9 (4.9)**</td>
<td>20.4 (13.4)**</td>
<td>44.8 (10.6)</td>
<td>11.8 (10.4)</td>
</tr>
<tr>
<td>Depression (BDI)</td>
<td>9.5 (3.0)</td>
<td>13.2 (8.2)</td>
<td>14.1 (7.5)</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>2.1 (3.3)**</td>
<td>9.0 (6.0)**</td>
<td>12.8 (7.0)</td>
<td>2.7 (3.7)</td>
</tr>
</tbody>
</table>

Unless stated otherwise, the groups "completely recovered" and "feeling much better" are combined into one group ("improved": $n = 50$, 20%) in further analyses, because of the small sample size of the first. There were no differences in any of the dimensions at initial assessment between patients who reported that complaints had remained the same and patients who reported that they had become worse. At follow up assessment, there were only differences between these two groups on the dimensions of functional impairment and sense of control over symptoms. As both groups seemed similar, they were combined into one group ("non-improved": $n = 196$, 80%).

**TREATMENT DURING FOLLOW UP INTERVAL**

During the follow up interval of 18 months, 143 (58%) patients had visited a specialist and 110 (45%) had visited an alternative practitioner. One hundred and seventy one (70%) patients had visited either of them. Of the patients that had visited a specialist, 78 were treated by that specialist (32% of the total sample). All patients that had visited an alternative practitioner received treatment. One hundred and forty six (59%) patients had been treated by either a specialist or an alternative practitioner during the follow up interval.

At initial assessment, 87% of patients used medication, mainly homeopathic drugs (38%), vitamins (26%), analgesics (24%), antidepressants (17%), antibiotics (14%), and sleep medication (7%). At follow up, 75% used medication (homeopathic drugs: 21%; vitamins: 20%; analgesics: 28%; antidepressants: 9%; antibiotics: 1%; sleep medication: 9%). At initial assessment, there was no significant difference between improved and non-improved patients in the proportion of patients using medication (82% $v$ 88%). At follow up, the proportion of patients using medication was significantly smaller in improved patients than in non-improved patients (61% $v$ 78%: $P < 0.01$).

**SELF REPORTED CHANGE COMPARED WITH DIMENSIONAL CHANGE AND IN SEVERITY OF FATIGUE**

Table 1 shows data on the seven outcome measures and depression at initial assessment and follow up assessment. These data are presented for recovered, improved, and non-improved patients and healthy subjects separately. At initial assessment all patient groups had higher fatigue severity scores than healthy subjects. Recovered and improved patients showed significant improvement in fatigue severity. At follow up, recovered patients had similar scores to healthy subjects, whereas improved patients still had higher scores than healthy subjects.

For the other outcome measures each patient group had significantly higher scores than healthy subjects at initial assessment, except for patients recovered at follow up, whose scores were similar to healthy subjects on social functioning and sleep disturbances. The recovered patients showed significant...
improvement on the outcome measures and at follow up had scores similar to healthy subjects on all dimensions. Improved patients showed significant improvement on psychological wellbeing, concentration problems, and subjective experience. However, at follow up these patients still had higher scores on these outcome measures than healthy subjects. Non-improved patients only significantly improved on psychological wellbeing. At follow up this group scored higher than healthy subjects on all outcome measures.

Analyses on difference scores showed that recovered and improved patients showed significantly greater improvement than non-improved patients on all outcome measures \( (P < 0.017) \). Recovered patients had greater improvement on all outcome measures, but this was only significant in fatigue severity and the subjective experience dimension \( (P < 0.017) \).

For depression, at initial assessment there were no significant differences between recovered, improved, and non-improved patients on the BDI. All groups had significantly higher scores than healthy subjects. At follow up, recovered and improved patients showed significant improvement. Compared with healthy subjects, at follow up, non-improved and improved patients still had higher BDI scores, but recovered patients had similar scores.

**PREDICTION OF SELF REPORTED CHANGE BY VARIABLES AT INITIAL ASSESSMENT**

Self reported change was not related to sociodemographic variables, medication, or being treated by a specialist or alternative practitioner. Self reported improvement was related to younger age, shorter duration of complaints, lower fatigue severity, and less abnormal scores in the dimensions sense of control over symptoms, functional impairment, physical attributions, and subjective experience (table 2).

Entering these variables in stepwise multiple regression analyses, self reported improvement was predicted by sense of control over symptoms, lower fatigue severity, and less abnormal scores in the dimensions psychological wellbeing, functional impairment, social functioning, sleep problems, concentration problems, sense of control over symptoms, and subjective experience (table 2). Treating these as independent variables in stepwise multiple regression analyses, high fatigue severity at follow up was predicted by three variables at initial assessment: high fatigue severity, lack of control over symptoms, and functional impairment (table 3). To evaluate the causal direction between fatigue severity and predictors, cross lagged panel analyses were performed (table 4). The correlation between sense of control over symptoms and fatigue severity at follow up was significantly stronger than the correlation between fatigue severity at initial assessment and sense of control over symptoms at follow up, which implies that the direction of causation goes from sense of control over symptoms to fatigue severity. The cross lagged panel analyses between fatigue severity and functional impairment was not significant.

**PREDICTION OF FATIGUE SEVERITY AT FOLLOW UP BY VARIABLES AT INITIAL ASSESSMENT**

High fatigue severity at follow up was related to the following variables at initial assessment: older age, high fatigue severity, and higher scores on the dimensions psychological wellbeing, functional impairment, social functioning, sleep problems, concentration problems, sense of control over symptoms, and subjective experience (table 2). Treating these as independent variables in stepwise multiple regression analyses, high fatigue severity at follow up was predicted by three variables at initial assessment: high fatigue severity, lack of control over symptoms, and functional impairment (table 3). To evaluate the causal direction between fatigue severity and predictors, cross lagged panel analyses were performed (table 4). The correlation between sense of control over symptoms and fatigue severity at follow up was significantly stronger than the correlation between fatigue severity at initial assessment and sense of control over symptoms at follow up, which implies that the direction of causation goes from sense of control over symptoms to fatigue severity. The cross lagged panel analyses between fatigue severity and functional impairment was not significant.

**Discussion**

As far as we know, this is the first prospective study on the natural course in chronic fatigue syndrome in which a comprehensive assessment of cognitive, emotional, behavioural and social functioning took place at initial assessment and at follow up.
In the present sample only a few patients recovered or improved and problems at work were considerable. Medical consumption was also high. All patients had visited specialists at some time before initial assessment. During the follow up interval, nearly all patients continued visiting specialists and alternative practitioners, and used medication.

Self reported improvement was in accordance with improvement on the seven outcome measures of clinical status and disability. Patients who had recovered at follow up showed significant improvement in five out of seven outcome measures. At follow up, their scores were similar to healthy subjects. Improved patients showed significant improvement on four out of seven outcome measures. However, they still had higher scores than healthy subjects on all outcome measures at follow up, indicating that recovery was not complete. The finding that on three out of seven outcome measures these patients did not show improvement underlines the importance of multidimensional assessment in studies on prognosis. Recovered and improved patients showed significantly greater improvement than non-improved patients. Although for all outcome measures improvement was greater in recovered patients than in improved patients, these differences were only significant for fatigue severity and the subjective experience dimension. The small number of recovered patients may have prevented the differences in the other five outcome measures from reaching significance.

Demographic variables were not predictive for self reported improvement; nor were receiving treatment by a specialist or an alternative practitioner. Improvement could be predicted by sense of control over symptoms, lower fatigue, shorter duration of complaints, and relative absence of physical attributions.

Psychological wellbeing (including depression) did not predict improvement in this study, although others have suggested that this factor plays a part in the perpetuation of complaints. Sharpe et al. found a relation between depression and improvement, but they measured depression retrospectively. In the study by Clark et al., diagnosis of depression at initial assessment was not a predictor for the course of the condition.

Avoidance of physical activity is also thought to play a part in the perpetuation of complaints, but the present study is not conclusive on this issue. The method of measuring avoidance of physical activity at initial assessment was crude, providing information only on whether or not a patient avoided physical activity, rather than how much or how often. At follow up assessment, avoidance of physical activity was scored on a 4 point scale (ranging from “seldom” to “almost all of the time”) to increase discriminative power. This variable correlated with self reported improvement ($r = 0.13$) and fatigue severity at follow up ($r = 0.21$). However, being assessed at follow up only, the causal direction of the relation with improvement and fatigue severity is unclear: a reduction in avoidance of physical activity may merely reflect improvement rather than causing it. Moreover, indicating how often one avoids physical activity does not give direct information on the actual level of the activity, and that is what really is of interest. We are currently performing studies in which physical activity (besides self report questionnaires) is assessed by a combination of self observation lists and a motion sensing device (the Actometer) during two week observation periods.

Patients with a short duration of complaints were more likely to improve than patients with a long duration. Because duration of complaints was long in the present sample (median 4.5 years), it might be speculated that most patients had already developed a chronic state. In that case improvement is less likely. Thus it is unclear whether present findings concerning improvement rate and predictors for improvement also apply to patients with a relatively short duration of complaints. A prospective study with a larger subset of patients with a short duration of complaints (six months to one year) seems warranted.

Attributing complaints to physical causes is one of the most distinctive features of patients with chronic fatigue syndrome and a high score on this dimension predicted poor prognosis. This agrees with Sharpe et al. who found that belief in a viral cause was related to functional impairment, but they assessed this variable at follow up only, thus leaving the question of causal direction unanswered. Like us, Wilson et al. found that the belief that a physical process explained all symptoms predicted poor outcome. Bonner et al. only found a trend for physical attributions to be associated with poor outcome. However, this study was small and really concerned the outcome of treated patients. Unfortunately, the study by Clark et al. did not investigate cognitive factors.

The role of the subjective sense of control over symptoms (self efficacy expectations) has not been investigated in a prospective study before. In the present study, sense of control over symptoms was a predictor of fatigue severity and was the strongest predictor of improvement. It might be speculated that a lack of subjective sense of control over symptoms is the result of severe fatigue. However, cross lagged panel analyses suggested that subjective sense of control over symptoms caused fatigue severity and not fatigue caused lack of sense of control. The present findings underline the role of attributions and cognitions in the perpetuation of the complaints.

The study of Clark et al. is probably the only other study on the natural course in which predictors were assessed prospectively. Like the present study they found 3% of patients reporting full recovery, but the improvement rate (38%) was higher than in
the present study (17%). There may be several reasons for this discrepancy. Firstly, the follow-up interval in the study by Clark et al. was substantially longer (mean 2.5 years), which may increase the chance of improvement. Secondly, on a subjective change scale ranging from -7 (significant worsening) to +7 (complete recovery) the authors used an arbitrary cut off point for improvement of more than +3. No validation of this cut off for improvement had been performed by comparing improved patients with healthy subjects as has been done in the present study.

There are methodological issues which should be considered. Some non-responders at initial assessment were more likely to go to school, but were less likely to work, and had more problematic scores on social functioning. However, on all other variables, notably all factors predicting improvement and fatigue severity, there were no differences between responders and non-responders. Therefore, we think that it is very unlikely that the follow-up sample is biased.

The present findings have therapeutic and theoretical implications. Several authors have proposed cognitive-behavioural therapy, which in general is directed at increasing physical activity and reducing depression to prevent chronicity. Our findings stress the role of cognitive processes in the perpetuation of complaints. We think that in therapeutic trials more attention should be paid to breaking down physical attributions and increasing the subjective sense of control over symptoms.


