

# The Longitudinal Outcome of Fibromyalgia: A Study of 1555 Patients

BRIAN WALITT, MARY-ANN FITZCHARLES, AFTON L. HASSETT, ROBERT S. KATZ, WINFRIED HÄUSER, and FREDERICK WOLFE

**ABSTRACT. Objective.** To describe the diagnosis status and outcome of patients diagnosed with fibromyalgia (FM) by US rheumatologists.

**Methods.** We assessed 1555 patients with FM with detailed outcome questionnaires during 11,006 semiannual observations for up to 11 years. At entry, all patients satisfied American College of Rheumatology preliminary 2010 FM criteria modified for survey research. We determined diagnosis status, rates of improvement, responder subgroups, and standardized mean differences (effect sizes) between start and study completion scores of global well-being, pain, sleep problems, and health related quality of life. (QOL)

**Results.** The 5-year improvement rates were pain 0.4 (95% CI 0.2, 0.5), fatigue 0.4 (95% CI 0.2, 0.05), and global 0.0 (95% CI -0.1, 0.1). The standardized mean differences were patient global 0.03 (95% CI -0.02, 0.08), pain 0.22 (95% CI 0.16, 0.28), sleep problems 0.20 (95% CI 0.14, 0.25), physical component summary of the Short-form 36 (SF-36) 0.11 (95% CI -0.14, -0.07), and SF-36 mental component summary 0.03 (95% CI -0.07, 0.02). Patients switched between criteria-positive and criteria-negative states, with 716 patients (44.0%) failing to meet criteria at least once during 4228.5 patient-years (7448 observations). About 10% of patients had substantial improvement and about 15% had moderate improvement of pain. Overall, FM severity worsened in 35.9% and pain in 38.6%.

**Conclusion.** Although we found no average clinically meaningful improvement in symptom severity overall, 25% had at least moderate improvement of pain over time. The result that emerged from this longitudinal study was one of generally continuing high levels of self-reported symptoms and distress for most patients, but a slight trend toward improvement. (J Rheumatol First Release July 15 2011; doi:10.3899/jrheum.110026)

*Key Indexing Terms:*  
FIBROMYALGIA

OUTCOME

IMPROVEMENT

From Georgetown University, Washington Hospital Center, Washington, DC; Montreal General Hospital, Division of Rheumatology, McGill University, Montreal, Quebec, Canada; Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, Michigan; Rush University Medical Center, Chicago, Illinois; Department of Psychosomatic Medicine and Psychotherapy, Technische Universität München, Munich, Germany; and the National Data Bank for Rheumatic Diseases and University of Kansas School of Medicine, Wichita, Kansas, USA.

Dr. Walitt has received consulting fees from Jazz Pharmaceuticals. Dr. Häuser has received honoraria from Eli-Lilly, Janssen-Cilag, Mundipharma, and Pfizer, and travel support from Eli Lilly. Dr. Hassett has received research funding from Bristol-Myers Squibb and Jazz Pharmaceuticals and is a consultant to Jazz Pharmaceuticals and Bristol-Myers Squibb. Dr. Fitzcharles has received consultant fees, speaking fees, and/or honoraria from Boehringer Ingelheim, Eli-Lilly, Paladin, Pfizer, and Purdue.

B. Walitt, MD, MPH, Georgetown University, Washington Hospital Center; M-A. Fitzcharles, MB, ChB, Montreal General Hospital, Division of Rheumatology, McGill University; A.L. Hassett, PsyD, Department of Anesthesiology, University of Michigan Medical School; R.S. Katz, MD, Rush University Medical Center; W. Häuser, MD, Department of Psychosomatic Medicine and Psychotherapy, Technische Universität München; F. Wolfe, MD, National Data Bank for Rheumatic Diseases, University of Kansas School of Medicine.

Address correspondence to Dr. F. Wolfe, National Data Bank for Rheumatic Diseases, 1035 N. Emporia, Suite 288, Wichita, KS 67214. E-mail: fwolfe@arthritis-research.org

Accepted for publication May 6, 2011.

The outcome of fibromyalgia (FM) — whether patients with the illness improve and to what extent they improve — has important implications that relate to clinical care and public policy with respect to diagnosis, disablement, and overall management. A number of studies, generally with samples  $\leq 100$ , have addressed the outcome and/or stability of FM outcomes<sup>1,2,3,4,5,6,7,8,9,10,11,12</sup>, and several studies have made similar assessments for persons with chronic widespread pain<sup>13,14,15</sup>, more common and less symptomatically severe than FM. As most of the FM studies were small and only 2 evaluated patients for more than one followup assessment, there remains limited knowledge of the stability or rate of change of symptoms over time. The consequence is that a picture of FM outcome has not clearly emerged.

We used the American College of Rheumatology (ACR) 2010 preliminary diagnostic criteria<sup>16</sup>, as modified for survey research<sup>17</sup>, to analyze prospectively collected data on 1555 patients with FM during 11,006 semiannual observations. We used the modified criteria because the preliminary diagnostic criteria are not suitable for survey research. The objectives of this study were to assess (1) the stability, rate, and overall change in FM symptoms over time; (2) the stability of FM criteria and the rate of change from criteria-

positive to criteria-negative status; and (3) the relation of symptoms to criteria, the predictors of symptoms, and criteria change.

## MATERIALS AND METHODS

**Patients and diagnoses.** Beginning in 1998, we assessed participants in the National Data Bank for Rheumatic Diseases (NDB) longitudinal study of FM outcomes<sup>18</sup>. Participants are volunteers, recruited from the practices of US rheumatologists, who complete mailed or Internet questionnaires at 6-month intervals (January and July). They are not compensated for their participation. The NDB utilizes an open-cohort design in which patients are enrolled continuously. The mean entry date for FM participants was July 2002. The diagnosis of FM was made by the patient's rheumatologist or confirmed by the patient's physician in cases that were self-referred (28%)<sup>18</sup>.

This study was approved by the Via Christi Regional Medical Center Institutional Review Board.

**Entry criteria.** For the purposes of participation in this study, patients were required to fulfil these criteria: (1) to have a rheumatologist diagnosis of FM prior to enrollment into the NDB; (2) to satisfy ACR 2010 diagnostic criteria modified for surveys and clinical studies at the time of the initial NDB assessment; and (3) to have completed  $\geq 2$  NDB semiannual questionnaires (Figure 1)<sup>17</sup>. Some of the modified 2010 criteria variables were

not available prior to 2009; therefore, we used the symptom intensity scale (SI)<sup>19</sup> to derive diagnostic criteria and the fibromyalgia severity (FS) scale for observations in this study prior to 2009. The FS scale (also known as the fibromyalginess scale)<sup>17,20</sup>, is similar to the SI scale<sup>21</sup>. Although the SI scale combines a visual analog scale (VAS) for fatigue with the Widespread Pain Index (WPI) and the FS scale combines the WPI with a 4-item symptom severity scale (not available in this study), the SI and FS scales are effectively the same in terms of performance. For when the SI scale is transformed to the same scale length as the FS (0 to 31), the Pearson correlation coefficient of the scales is 0.963 and Lin's concordance coefficient is 0.956<sup>22</sup>. This indicates that the scales have almost exactly the same performance characteristics. We have reported elsewhere that the "scale captures well the essential content of FM or what we have called 'fibromyalginess'"<sup>17,20</sup>. A cutpoint of 13 best separates persons with and without FM, using the modified diagnostic criteria, when unselected patients with rheumatic disease are studied<sup>17</sup>. The modification of the 2010 criteria for diagnosis with data prior to 2009 in the NDB was: (WPI  $\geq 7$  and VAS fatigue  $> 5$ ) or (WPI  $\geq 7$  and VAS fatigue  $\leq 5$ ) and a count of somatic symptoms ( $\geq 13$ ).

Of the 1555 participants who met entry criteria, 19% had data for 1 year (2 observations), 25% had data for 2 years (3–4 observations), and 56% had data for between 3 and 11.5 years (6–23 observations). The mean duration in the study was 4.0 (SD 2.9) years. As enrollment into the NDB was continuous, beginning in 1998, duration of followup reflects both year of entry

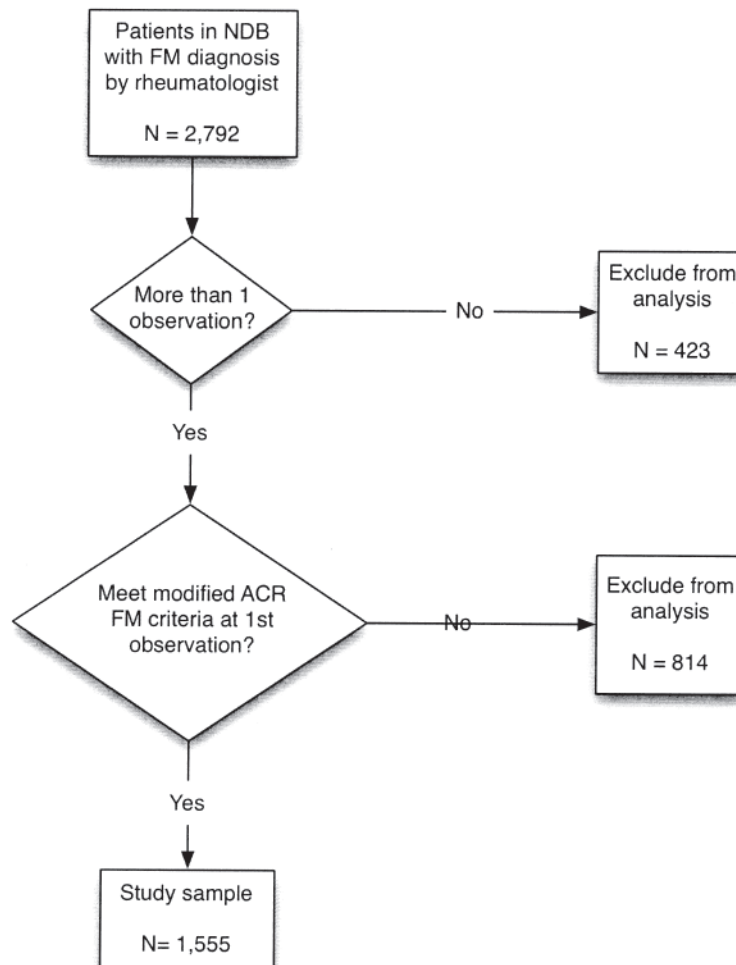


Figure 1. The process of patient enrollment and selection.

and withdrawal of consent. We also identified an additional 814 patients, diagnosed with FM by physicians but not meeting the modified ACR 2010 criteria at their first observation (Figure 1). These patients were excluded from the current study. However, at some time during followup 496 (60.9%) met the modified ACR criteria. A further 423 patients met study criteria but had only one observation and were excluded from the analysis (Figure 1). These 423 differed slightly from the 1555 valid participants at first assessment as follows (mean scores; 1555 in the included group, 423 the excluded group): FS scale (22.7, 23.5), VAS pain (6.8, 7.0), fatigue (7.6, 8.0), FM duration (13.7, 12.8), and age (52.8, 51.7).

**Study outcome variables.** The major study variables are shown in Tables 1 and 2. Patients completed the Short-form 36 (SF-36) version 1 from which the physical component summary (PCS) and the mental component summary (MCS) scores were calculated<sup>23,24</sup>. The primary time period of the SF-36 questionnaire was 4 weeks.

The SF-36 mental health subscale was normalized to a 0–10 mood scale, higher numbers indicating worse mental health, and this constituted the mood scale for the study. We also used the Euroqol (EQ-5D) to estimate a preference-based single measure of health status<sup>25</sup>. Lower scores represent worse outcome for the PCS, MCS, and EQ-5D. The FS scale was used to define FM severity (fibromyalginess). As noted above, the FS ranges from 0 to 31, a score of 13 considered the best dividing point between FM and non-FM states<sup>17</sup>.

To measure functional status, we used the Health Assessment Questionnaire disability index (HAQ)<sup>26</sup>. Fatigue, disturbed sleep, and global severity and pain were assessed by 0–10 visual analog scales. Global severity referred to “all the ways your illness affects you.” Body mass index (BMI) categories of underweight (BMI < 18.5 kg/m<sup>2</sup>), normal weight (BMI 18.5–24.9 kg/m<sup>2</sup>), overweight (BMI 25.0–29.9 kg/m<sup>2</sup>), and obesity (BMI 30.0 kg/m<sup>2</sup>) were determined according to the World Health Organization guidelines<sup>27</sup>. Patients reported all medication used within the previous 6 months on each questionnaire. Medications were classified as psychotrop-

ic (including antidepressants, anxiolytics, anticonvulsants, and similar medications), analgesic and nonsteroidal antiinflammatory drugs (NSAID), and/or opioids.

A count of somatic symptoms (0–37), similar to those reported in the ACR 2010 diagnostic criteria study<sup>16</sup>, was obtained. Depression within the last 6 months and work disability were determined by self-report.

**Statistical methods.** Using 11,006 observations, patients were studied during 6251 patient-years of followup. We used Kaplan-Meier life table methods and Cox regression to determine the risk of not satisfying FM criteria. To determine the rate of change in variables over time, we used generalized estimating equations (GEE), clustered on individual patients and adjusted for sex, baseline age, and baseline FM duration. Because all patients were not represented in all observations, GEE data are potentially biased. To be sure the results presented were substantially correct, we conducted a series of sensitivity analyses to determine if missing data biased the results. We used a series of qualitative and quantitative graphic and regression methods including fractional polynomial regression, restricted cubic splines, and line smoothers. In these analyses we compared data at different time periods for patients with and without generally complete data. We also used markers for dropouts (attrition) and short time in study; these variables were not significant in any model. The 1073 patients who dropped out of the study prematurely through death (4%) or withdrawal of consent, representing an attrition rate of 11% per semiannual questionnaire, differed minimally from the 482 who continued: (n = 482, n = 1073): FS scale (22.4, 22.9), VAS pain (6.6, 6.9), fatigue (7.5, 7.7), FM duration (13.8, 13.7), and age (52.6, 53.0). Overall, our analyses suggested that the data as presented are a fair representation of the course of FM.

We determined the standardized mean difference between variables at the first and final observation by dividing the mean difference by the pooled standard deviation. Confidence intervals were based on 100 bootstrap repetitions. As entry criteria constrained the variance of FM severity, WPI, fatigue and symptom count, the effect size for these variables is biased and was not reported. We used Cohen’s categories to categorize the magnitude of the effect size, with values > 0.2 indicating a small effect size, ≥ 0.5 a medium effect, and ≥ 0.8 a large effect<sup>28</sup>, and we used IMPACT recommendations to form and describe improvement groups<sup>29</sup>.

We used the Gonen and Heller K statistic for concordance probability to evaluate the discriminatory power and the predictive accuracy in the Cox proportional hazards analyses<sup>30</sup>. The K statistic may be interpreted similarly to an area under the receiver operating curve statistic.

To put study scores into a larger perspective, we also report scores from a random observation from 15,777 patients with rheumatoid arthritis (RA) participating in the NDB outcome study<sup>18</sup>. Approximately 22% of these patients satisfied survey criteria for FM. All analyses were performed using Stata, version 11.1<sup>31</sup>. Statistical significance was set at 0.05. All tests were 2-tailed.

## RESULTS

**Baseline characteristics.** Participants were almost all non-Hispanic White (92.5%) women (96.4%). The median household income was \$35,000. Between 29.1% and 34.5% were work-disabled. Pharmacologic FM-related treatments were used almost universally (96.8%), with 85.4% reporting use of analgesics, opioids, and NSAID and 86.9% using other drugs that included antidepressants, anti-anxiety, muscle relaxants, sleeping medications, and similar treatments. Current depression (within last 6 months) was reported by 47.0% (Table 1).

As shown in Table 2, clinical status variables were markedly abnormal at entry, indicating high levels of polysymptomatic distress. Subsequent analyses (below) report changes in these variables over the time of the study.

Table 1. Demographic and treatment characteristics at first observation.

Variable	Baseline Value, mean (SD) or %
Age, yrs	52.9 (11.4)
Sex, % female	96.4
Non-hispanic White, %	92.5
Education, %	
< 12 yrs	6.3
12–15 yrs	67.1
College graduate	26.7
Body mass index (BMI)	30.6 (7.6)
BMI WHO classification: overweight or obese, %	61.9
Smoking: current, %	15.5
Smoking: past, %	21.8
Social security disability (age ≤ 62 years)	29.1
Work disabled — self-report, %	34.5
Median total household income (\$US)	35,000
FM treatment, %	
NSAID or non-opioid analgesics	67.4
Opiate use	40.1
Any analgesic, opioid, or NSAID	85.4
FM pharmacological treatment: not analgesics, opioids, or NSAID	86.9
Any FM pharmacological treatment	96.8
Duration of fibromyalgia, yrs	13.7 (10.7)
Current depression (within last 6 months)	47.0

WHO: World Health Organization BMI categories; NSAID: nonsteroidal antiinflammatory drugs.

Table 2. Baseline and longitudinal improvement values for 1555 patients with fibromyalgia.

Variable	Baseline, mean (SD)	All Observations (N = 11,006) Within-patient SD	Estimated 5-year Improvement (95% CI)*	Mean (SD) of RA Patients (N = 15,777)
FM severity (0–31)	22.7 (4.8)	4.0	1.8 (1.5, 2.1)	11.2 (7.5)
Widespread Pain Index (0–19)	13.6 (5.7)	3.3	1.3 (1.0, 1.5)	5.7 (5.1)
Patient global (0–10)	5.7 (2.2)	1.5	0.0 (–0.1, 0.1)	3.7 (2.5)
Pain (0–10)	6.8 (2.1)	1.6	0.4 (0.2, 0.5)	3.8 (2.8)
Fatigue (0–10)	7.6 (1.9)	1.5	0.4 (0.3, 0.5)	4.4 (3.0)
Sleep disturbance (0–10)	6.8 (2.7)	1.9	0.4 (0.3, 0.6)	3.8 (3.1)
Symptom count (0–37)	16.4 (5.7)	3.4	0.8 (0.6, 1.1)	8.0 (6.2)
Mood (0–10)	4.5 (1.9)	1.0	0.5 (0.4, 0.6)	2.8 (1.9)
HAQ (0–3)	1.3 (0.6)	4.4	0.2 (0.00, 0.04)	1.0 (0.7)
PCS (0–100)	29.0 (7.7)	6.8	0.5 (0.1, 0.8)	36.6 (11.2)
MCS (0–100)	39.7 (11.6)	0.1	0.6 (0.1, 1.1)	49.0 (11.4)
EQ-5D (0–1)	0.57 (0.22)	4.0	0.03 (0.02, 0.04)	0.73 (0.19)

\* Based on generalized estimating equation analyses adjusted for baseline fibromyalgia duration, age, and sex. HAQ: Health Assessment Questionnaire disability index; PCS: physical component summary score; MCS: mental component summary score; EQ-5D: EuroQol.

*Improvement in FM symptoms over time.* We assessed change in FM symptom severity by comparing data at the first and last observations over a mean study duration of 4 years, and also by fitting longitudinal GEE models that utilized all of the 11,006 observations over the 11 years of followup. When the cohort was considered as a whole, there was little change in FM symptom severity. Figure 2 shows only slight to no changes in regard to the WPI, symptom count, fatigue, pain, mood, or HAQ scores between study

initiation and completion. Figure 3 demonstrates a small degree of improvement in FS scores over time, with improvement occurring more quickly in the earlier years. To describe the rate of change over time, we present estimated changes per 5 years rather than annually, using all (11,006) study observations (Table 2). The mean FS score was 22.7 (4.8) at study onset had an estimated 5-year improvement of 1.8 (1.5, 2.1) units, or 0.36 standardized units. The magnitude of this change is shown in Figure 3, right panel. Table

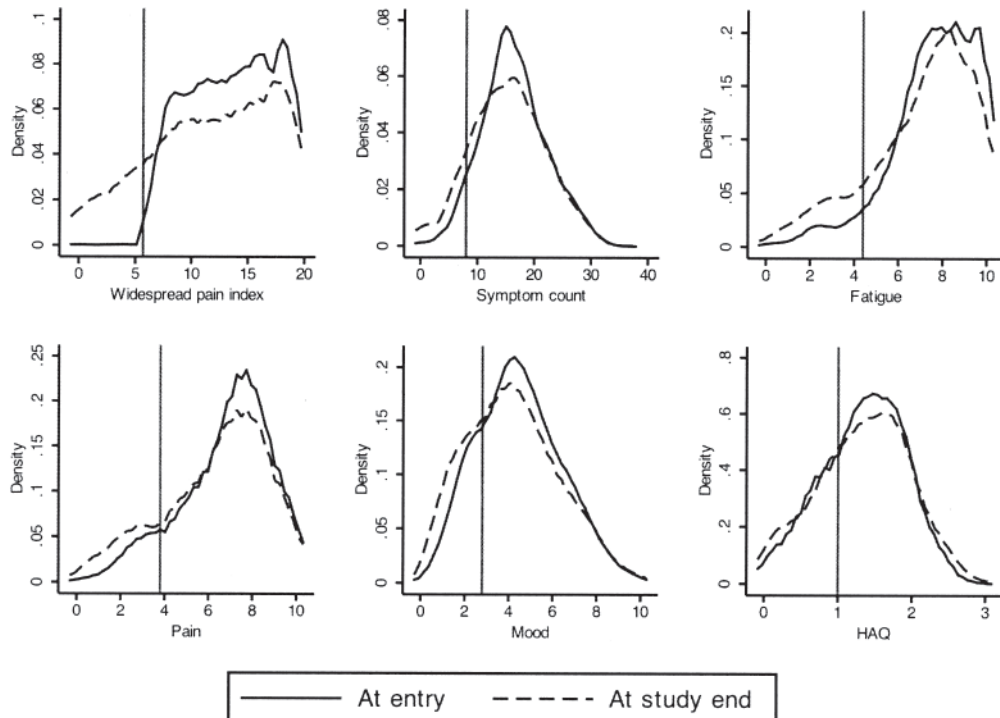


Figure 2. Key fibromyalgia symptoms scores at baseline and at study completion. Vertical lines indicate mean values in patients with rheumatoid arthritis, as shown in Table 1.

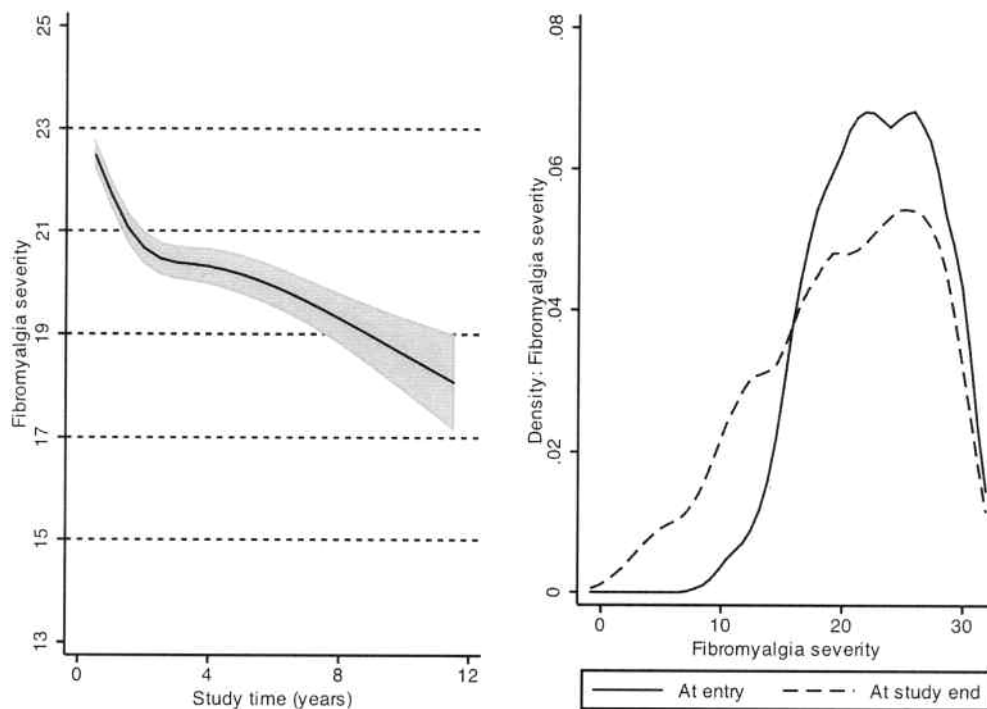


Figure 3. Changes in fibromyalgia severity score over time and between baseline and study completion. The left panel is based on generalized estimating equation analyses of all observations; the right panel contrasts first and last observations.

2 also demonstrates that slight improvements were noted for some important FM and QOL variables but not others. For example, the 5-year improvement for pain, fatigue, and sleep was 0.4 units, but there was no improvement in patient global severity. We used baseline FM duration as a covariate in the Table 2 analyses (column 2), and significant changes in 5-year improvement for each 1-year increase in the duration of FM were noted only for PCS worsening  $-0.04$  (95% CI  $-0.08, -0.00$ ), MCS improvement  $0.05$  (95% CI  $0.00, 0.10$ ), and mood improvement  $-0.01$  (95% CI  $-0.02, -0.00$ ). While statistically significant, these changes are small and should be considered clinically meaningless. In addition, there was considerable within-patient variability, as shown in Table 2 (column 2). The effect sizes for pain ( $0.22$ ; 95% CI  $0.12-0.28$ ) and sleep disturbance ( $0.20$ ; 95% CI  $0.16-0.25$ ) were small; the effect sizes on PCS, MCS, EQ-5D, and patient global were not substantial (Table 3, column 2).

Even so, Figure 3 (right panel) shows that some patients improved substantially. To explore this, we further characterized the degree of change for groups of patients by measuring response at the last observation. Table 3 shows that 10.2% of patients had a substantial response ( $\geq 50\%$ ) as measured by the FM severity variable. Among these responders, only 5.1% still satisfied FM criteria and showed only mild abnormalities in the other study variables. A further 13.6% had a moderate response, with clinically important improvement, although 43% still satisfied the criteria. By

contrast, 53.5% had no response, with variable scores worse than the baseline scores of all patients. Overall, FM severity worsened in 35.9% and pain in 38.6%. Other study variables generally demonstrated similar outcomes. Of interest, there was no difference in the duration of FM between groups at study closure: 17.2, 17.2, 17.7, and 16.5 years for the none, minimal, moderate, and substantial groups, respectively.

*Losing FM diagnosis: becoming criteria-negative.* Patients switched between criteria-positive and criteria-negative states. Seven hundred sixteen patients (44.0%) failed to meet criteria at least once during 4228.5 patient-years (7448 observations). The incident rate for failing to meet criteria ever during the study was 16.9 (15.7, 18.2) per 100 patient-years. To determine how many participants had not met FM criteria at study closure, we repeated the analysis considering only each participant's most recent (last) observation. In this analysis, 378 patients (24.3%) failed to meet criteria, for an incident rate of 6.0 (5.5, 6.6) per 100 patient-years. However, less than half the participants who failed to meet criteria at one observation continued to not meet criteria at the next observation.

As 44.0% of patients initially criteria-positive failed to meet criteria at some time during the study, we studied the FM severity characteristics of criteria-positive observations compared with criteria-negative observations to determine the benefit that occurs with the criteria-negative state. Table 4A examines all 11,006 study observations for the 1555

Table 3. Baseline, standardized mean difference, and percentage improvement values for 1555 patients with fibromyalgia.

Outcome Measures	Standardized Mean Difference (95% CI)*	All Patients, Baseline, N/%	Degree of Improvement Compared with Baseline**			
			None, < 10%, N/%	Minimal, 10–29%, N/%	Moderate, 30–49%, N/%	Substantial, ≥ 50%, N/%
FM severity (0–31)		1555/100	832/53.5	353/22.7	212/13.6	158/10.2
Mean	NV	22.7	24.3	19.7	14.0	7.6
FM criteria (+)		1555/100	832/94.5	353/82.4	92/43.4	158/5.1
WPI (0–19)	NV	1555/100	806/51.8	292/18.8	176/11.3	281/18.1
Mean		13.6	15.4	11.6	8.6	3.3
Patient global (0–10)	0.03 (–0.02, 0.08)	1555/100	933/60.0	284/18.3	164/10.5	174/11.2
Mean		5.7	6.6	5.6	4.0	2.1
Pain (0–10)	0.22 (0.16, 0.28)	1555/100	887/57.0	324/20.8	151/9.7	193/12.4
Mean		6.8	7.5	6.3	4.5	2.1
Fatigue (0–10)	NV	1555/100	866/55.6	384/24.7	142/9.1	163/10.5
Mean		7.6	8.2	7.0	4.8	2.3
Sleep disturbance (0–10)	0.20 (0.14, 0.25)	1489/100	836/56.1	278/18.7	128/8.6	247/16.6
Mean		6.8	7.8	6.6	4.7	1.7
Symptom count (0–37)	NV	1544/100	874/56.2	367/23.6	169/10.9	144/9.3
Mean		16.4	18.3	14.7	10.9	4.7
Mood (0–10)	0.15 (0.10, 0.20)	1548/100	827/53.2	330/21.3	196/12.6	195/12.5
Mean		4.5	5.1	4.0	3.0	1.5
HAQ (0–3)	0.00 (–0.03, 0.03)	1555/100	994/63.9	289/18.6	135/8.7	137/8.8
Mean		1.3	1.6	1.3	0.9	0.3
PCS (0–100)	–0.11 (–0.14, –0.07)	1555/100	975/62.7	338/21.7	152/9.8	90/5.8
Mean		29.0	27.6	33.3	34.9	36.5
MCS (0–100)	–0.03 (–0.07, 0.02)	1555/100	987/63.5	289/18.6	143/9.2	136/8.7
Mean		39.7	36.8	43.7	45.5	46.1
EQ-5D (0–1)	–0.10 (–0.16, –0.05)	1208/100	821/68.0	135/11.2	20/1.7	232/19.2
Mean		0.57	0.55	0.62	0.58	0.65

\* Standardized mean difference (effect size) = last observation mean minus first observation mean divided by pooled standard deviation. Positive values represent improvement except for PCS, MCS, and EQ-5D, where negative values represent improvement. \*\* Degree of improvement = percentage improvement at last observation compared with baseline observation. NV: not valid; HAQ: Health Assessment Questionnaire disability index; PCS and MCS: physical and mental component scores from SF-36 (Short-form 36). WPI: Widespread Pain Index.

patients. Table 4B restricts the analyses to 1177 patients (7828 observations) who were criteria-positive at their last observation. Table 4B gives insight into patients who were positive at study start and completion, while Table 4A lifts the restriction of being criteria-positive at the final observation. Taken together, these tables show that there were generally small differences between criteria-positive and negative states. Even when patients become criteria-negative they retain abnormal scores, with little change for many variables. Of particular interest, the PCS and EQ-5D scores remain very abnormal, and the symptom count remains high.

*Predictors of becoming criteria-negative.* We examined the study variables for their ability to predict conversion from the initial criteria-positive state to a criteria-negative state using Cox regression. Gonen and Heller's K statistics were as follows: 0.51 marital status, antidepressant use; 0.52 smoking, current depression, MCS, sleep disturbance; 0.53 NSAID use; 0.54 BMI, opioid use, education level; 0.55 work disability, mood; 0.56 comorbidity; 0.57 symptom count, fatigue, global severity; 0.58 PCS; 0.59 EQ-5D, HAQ; 0.62 WPI; and 0.64 FM severity. These results indicate that demographic and treatment variables poorly pre-

dicted change in criteria status. However, FM severity and WPI components of the FM criteria were the best predictor variables. Patients with more abnormal scores, particularly the FM severity score, were less likely to convert.

## DISCUSSION

The picture that emerges from our longitudinal study of FM is one of continued high levels of self-reported symptoms and distress for the cohort as a whole. This is the first report to document waxing and waning of symptoms over time, with patients moving in and out of positive FM status. Through followup that extended to 11 years, we observed that symptom severity changed little, although the overall trend was for improvement. The changes in symptoms were very slight, with only a small effect seen when comparing start and ending scores of pain and sleep variables.

We also summarized the levels and changes in FM symptoms using the FM severity scale. This scale identifies the main content of the FM case definition and is derived from the ACR 2010 preliminary diagnostic criteria for FM and symptom severity<sup>16</sup>. FS changes over time, but, on average, the change is slight and the score remained very high.

Table 4A. Estimated mean values of study variables during criteria-positive and criteria-negative observations for all patients (N = 1555).

Variable	Criteria (+) Observations, mean (95% CI)	Criteria (-) Observations, mean (95% CI)	Difference, mean (95% CI)
Fibromyalgia severity (0–31)	22.5 (22.3, 22.7)	14.4 (14.1, 14.7)	8.1 (7.8, 8.4)
Widespread Pain Index (0–17)	13.5 (13.4, 13.7)	7.5 (7.3, 7.8)	6.0 (5.7, 6.3)
Patient global (0–10)	5.7 (5.6, 5.8)	4.9 (4.8, 5.0)	0.8 (0.7, 0.9)
Pain (0–10)	6.7 (6.6, 6.8)	5.4 (5.2, 5.5)	1.3 (1.2, 1.4)
Fatigue (0–10)	7.5 (7.5, 7.6)	5.7 (5.6, 5.9)	1.8 (1.7, 1.9)
Sleep disturbance (0–10)	6.6 (6.5, 6.7)	5.5 (5.3, 5.6)	1.1 (1.0, 1.2)
Symptom count (0–37)	16.4 (16.1, 16.7)	13.1 (12.8, 13.4)	3.3 (3.0, 3.6)
Mood (0–10)	4.3 (4.2, 4.4)	3.8 (3.7, 3.9)	0.5 (0.4, 0.6)
HAQ (0–3)	1.4 (1.3, 1.4)	1.2 (1.2, 1.3)	0.1 (0.1, 0.2)
PCS (0–100)	29.3 (28.9, 29.6)	31.7 (31.2, 32.2)	-2.5 (-2.8, -2.1)
MCS (0–100)	39.5 (39.0, 40.0)	42.5 (41.8, 43.1)	-3.0 (-3.5, -2.5)
EQ-5D (0–1)	0.56 (0.55, 0.57)	0.66 (0.62, 0.64)	-0.07 (-0.08, -0.06)

HAQ: Health Assessment Questionnaire disability index; PCS and MCS: physical and mental component scores from SF-36 (Short-form 36).

Table 4B. Estimated mean values of study variables during criteria-positive and criteria-negative observations for 1177 patients who were criteria-positive at their last observation.

Variable	Criteria (+) Observations, mean (95% CI)	Criteria (-) Observations, mean (95% CI)	Difference, mean (95% CI)
Fibromyalgia severity (0–31)	23.3 (23.1, 23.5)	15.4 (15.0, 15.7)	7.9 (7.5, 8.3)
Widespread Pain Index (0–17)	14.1 (13.9, 14.2)	8.1 (7.7, 8.5)	5.9 (5.5, 6.3)
Patient global (0–10)	5.9 (5.9, 6.0)	5.1 (4.9, 5.3)	0.9 (0.7, 1.0)
Pain (0–10)	6.9 (6.8, 7.0)	5.6 (5.4, 5.8)	1.2 (1.1, 1.4)
Fatigue (0–10)	7.7 (7.6, 7.8)	6.0 (5.8, 6.2)	1.7 (1.5, 1.9)
Sleep disturbance (0–10)	6.7 (6.7, 6.9)	5.9 (5.7, 6.1)	0.9 (0.7, 1.1)
Symptom count (0–37)	17.2 (16.9, 17.4)	14.3 (13.9, 14.7)	2.8 (2.5, 3.2)
Mood (0–10)	4.4 (4.3, 4.5)	4.0 (3.9, 4.2)	0.4 (0.3, 0.5)
HAQ (0–3)	1.4 (1.4, 1.4)	1.3 (1.2, 1.3)	0.1 (0.1, 0.2)
PCS (0–100)	28.3 (28.0, 28.7)	30.5 (29.9, 31.2)	-2.2 (-2.7, -1.6)
MCS (0–100)	38.7 (38.1, 39.2)	41.5 (40.6, 42.4)	-2.8 (-3.5, -2.1)
EQ-5D (0–1)	0.53 (0.53, 0.55)	0.60 (0.59, 0.62)	-0.06 (-0.08, -0.05)

HAQ: Health Assessment Questionnaire disability index; PCS and MCS: physical and mental component scores from SF-36 (Short-form 36).

Even if the mean symptom scores were stable, individual patients' scores varied considerably over the study, and almost half of each variable's variance was explained by the within-patient variance (data not shown). Thus at the patient level, visit-to-visit scores often reflected clinically significant change. In addition, at the final observation we noted that around 10% of patients had substantial improvement, with levels of symptom variables close to minimal. One might speculate whether this group of patients were free of FM and FM symptoms or whether this outcome represents a transitory state, given the fluctuations of symptom severity within patients and the long duration of FM (16.5 years) in this group. When the results of this group are combined with those having moderate improvement, it appears that approx-

imately 25% of patients could be considered to have a good outcome. Thus, within the cohort where the average improvement was small, there is a subgroup of 25% with meaningful improvement, including 10% with substantial improvement.

Given the within-patient variability of symptoms, it might be expected that patients might shift from criteria positivity to criteria negativity. We found exactly that. Over the course of the study, 716 patients (44.0%) failed to meet criteria at least once, resulting in an incidence rate of 16.9 (15.7, 18.2) events per 100 patient-years. At the last study observation 24.3% failed to meet criteria, for an incidence rate of 6.0 (5.5, 6.6) events per 100 patient-years.

In addition, more than half the participants who failed to

meet criteria at one observation met criteria again at the next observation. Given the relative stability of symptom severity noted, the data suggest to us an apparent disparity between “no longer satisfying criteria” and level of symptoms. That is, it is much easier to become FM criteria-negative than it is to improve substantially. Given this observation, we suggest that symptom levels are superior to and provide more information than criteria status; and we recommend that symptom levels, perhaps through the use of the FS scale, be used to clinically define patients with FM over time. Symptom levels are not yoked to a particular cutpoint, as required by current criteria. That is, patients can be clinically defined in relation to prior and future evaluations rather than by criteria status at a single point in time. As an example of such a situation, consider the status of pain patients with a PCS of 35, surely very symptomatic, but not generally satisfying current criteria. The use of symptom scales rather than criteria allows such patients to be evaluated as part of a continuum of biopsychosocial distress<sup>32</sup>.

The outcome of FM has been the subject of a number of reports in usually small studies, utilizing different criteria and assessments over short periods of time, and with sometimes conflicting results<sup>1,2,3,4,5,6,7,8,9,10,11</sup>. However, in general, the results of these studies tend to suggest little change in symptoms. A study that is particularly germane to our current study is the careful report of Fitzcharles, *et al*<sup>12</sup> that collected similar variables. Patients in their study of 70 patients (9 patients did not fulfill 1990 ACR criteria at entry) improved over 3 years, with effect sizes for pain and patient global of approximately 0.48 and 0.37, respectively, compared with 0.22 and 0.03 in our current study. The final values for pain 5.4, sleep disturbance 5.3, fatigue 5.7, global 5.5, and HAQ 0.94 were different from the values we obtained at entry of pain 6.8, sleep disturbance 6.8, fatigue 7.6, global 5.7, and HAQ 1.3, although the duration of FM was similar in both studies. However, the values of the current study were similar to those obtained in the ACR 2010 diagnostic criteria study of 196 patients: pain 6.5, sleep disturbance 6.5, fatigue 7.0, and HAQ 1.3<sup>16</sup>. These data suggest the results of the current study may be more generally correct; but it seems possible the careful clinical care provided by Fitzcharles, *et al*<sup>12</sup> could be responsible for the improved outcome.

These results raise a number of important points: How can it be that most patients improved only minimally despite treatment, and continued to have very high levels of symptoms? In considering the possibility that patients in this study were systematically different from fibromyalgia patients in other settings, we have observed that FM symptom levels were similar between the current study and the clinical trial of Arnold, *et al*<sup>33</sup>. Bradley, *et al* suggest that a small fraction of persons with FM are “non-patients”<sup>34</sup>, but non-patients are just that, non-patients, and it is not clear

that such a group would satisfy the modified ACR 2010 criteria that are highly dependent on symptom severity.

Related to the above discussion is the possibility that our inclusion criteria resulted in the exclusion of mild patients. In preparing the sample for this study, we excluded 34.4% of patients who had been diagnosed with FM by a physician, but did not satisfy our modified ACR 2010 criteria at entry. This is similar to the percentage with a prior diagnosis of FM who did not satisfy ACR 1990 criteria when reexamined for the 2010 criteria study<sup>16</sup>. It seems likely that patients with “near FM” who did not satisfy entry criteria might have satisfied them if followed long enough. In fact, of the 814 patients who did not satisfy entry criteria, 60.1% satisfied criteria during followup. Therefore, we believe that the use of binary criteria for a continuous symptom disorder is problematic. These observations reinforce the concept that a person may have an established diagnosis of FM, but with fluctuations in symptom severity over time.

Although our study was not designed to examine treatment efficacy, our report represents the outcome of a large series of FM patients who received specialist care for the disorder from US rheumatologists. While it seems possible that there are better treatments<sup>35,36</sup> or that some patients received suboptimal care, we have no data to address such issues: our results represent outcome for care as it was delivered. While observational studies are suboptimal for determining efficacy, they are able to provide accurate estimates of the level of symptoms in the community. The data suggest that one should be wary of attributing substantial clinical benefit to currently available pharmacological therapies, at least as they are used in the community in the long term.

Some limitations should be noted. Our population was self-selected. Participants in survey research are usually better educated and have better outcomes than nonparticipants<sup>37</sup>. On the other hand, it is possible that patients have self-selected for chronicity. Men were underrepresented in the sample (3.6%), but were similar to F. Wolfe’s practice (7.9%)<sup>37</sup>.

Despite drawbacks, this report describes a very large prospective study of the clinical course of FM. Over a mean time course of 4 years, we were unable to appreciate a clinically important average change in FM symptom severity over time. However, approximately 20%–25% of the patients reported at least a moderate improvement. These data provide clinicians and patients with realistic expectations on the course of FM in routine clinical care.

## REFERENCES

1. Forseth KO, Forre O, Gran JT. A 5.5 year prospective study of self-reported musculoskeletal pain and of fibromyalgia in a female population: Significance and natural history. *Clin Rheumatol* 1999;18:114-21.
2. Felson DT, Goldenberg DL. The natural history of fibromyalgia. *Arthritis Rheum* 1986;29:1522-6.
3. Hawley DJ, Wolfe F, Cathey MA. Pain, functional disability, and



- psychological status: a 12-month study of severity in fibromyalgia. *J Rheumatol* 1988;15:1551-6.
4. Ledingham J, Doherty S, Doherty M. Primary fibromyalgia syndrome — an outcome study. *Br J Rheumatol* 1993;32:139-42.
  5. Granges G, Zilko P, Littlejohn GO. Fibromyalgia syndrome: assessment of the severity of the condition 2 years after diagnosis. *J Rheumatol* 1994;21:523-9.
  6. Kennedy M, Felson DT. A prospective long-term study of fibromyalgia syndrome. *Arthritis Rheum* 1996;39:682-5.
  7. Wigers SH. Fibromyalgia outcome: the predictive values of symptom duration, physical activity, disability pension, and critical life events — a 4.5 year prospective study. *J Psychosom Res* 1996;41:235-43.
  8. Bengtsson A, Backman E, Lindblom B, Skogh T. Long-term follow-up of fibromyalgia patients: clinical symptoms, muscular function, laboratory tests — an eight year comparison study. *J Musculoskel Pain* 1994;2:67-80.
  9. Baumgartner E, Finckh A, Cedraschi C, Vischer TL. A six year prospective study of a cohort of patients with fibromyalgia. *Ann Rheum Dis* 2002;61:644-5.
  10. White KP, Nielson WR, Harth M, Ostbye T, Speechley M. Does the label “fibromyalgia” alter health status, function, and health service utilization? A prospective, within-group comparison in a community cohort of adults with chronic widespread pain. *Arthritis Rheum* 2002;47:260-5.
  11. Nöller V, Sprott H. Prospective epidemiological observations on the course of the disease in fibromyalgia patients. *J Negative Results BioMedicine* 2003;2:4.
  12. Fitzcharles MA, Costa DD, Poyhia R. A study of standard care in fibromyalgia syndrome: a favorable outcome. *J Rheumatol* 2003;30:154-9.
  13. Andersson HI, Ejlertsson G, Leden I, Rosenberg C. Characteristics of subjects with chronic pain, in relation to local and widespread pain report — A prospective study of symptoms, clinical findings and blood tests in subgroups of a geographically defined population. *Scand J Rheumatol* 1996;25:146-54.
  14. Papageorgiou A, Silman A, Macfarlane G. Chronic widespread pain in the population: a seven year follow up study. *Br Med J* 2002;61:1071.
  15. Bergman S, Herrstrom P, Jacobsson LT, Petersson IF. Chronic widespread pain: a three year followup of pain distribution and risk factors. *J Rheumatol* 2002;29:818-25.
  16. Wolfe F, Clauw D, Fitzcharles MA, Goldenberg D, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care Res* 2010;62:600-10.
  17. Wolfe F, Clauw D, Fitzcharles MA, Goldenberg D, Häuser W, Katz RS, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: A modification of the ACR preliminary diagnostic criteria for fibromyalgia. *J Rheumatol* 2011;38 [in press].
  18. Wolfe F, Michaud K. The National Data Bank for Rheumatic Diseases: a multi-registry rheumatic disease data bank. *Rheumatology* 2011;50:16-24.
  19. Katz RS, Wolfe F, Michaud K. Fibromyalgia diagnosis: A comparison of clinical, survey, and American College of Rheumatology criteria. *Arthritis Rheum* 2006;54:169-76.
  20. Wolfe F. New criteria for fibromyalgia: a twenty year journey. *Arthritis Care Res* 2010;62:583-4.
  21. Wolfe F, Rasker JJ. The Symptom Intensity Scale, fibromyalgia, and the meaning of fibromyalgia-like symptoms. *J Rheumatol* 2006;33:2291-9.
  22. Lin LK. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-68.
  23. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
  24. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40-66.
  25. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol. York: Publications Unit, Centre for Health Economics, University of York; 1996.
  26. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
  27. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of WHO Expert Consultation. Geneva: World Health Organization; 1995.
  28. Cohen J. Statistical power analysis for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
  29. Dworkin R, Turk D, Wyrwich K, Beaton D, Cleeland C, Farrar J, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105-21.
  30. Gonen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 2005;92:965.
  31. Stata statistical software: Release 11.0. College Station, TX: Stata Corp.; 2009.
  32. Hauser W, Schmutzler G, Brahler E, Glaesmer. A cluster within the continuum of biopsychosocial distress can be labeled “fibromyalgia syndrome” — Evidence from a representative German population survey. *J Rheumatol* 2009;36:2806-12.
  33. Arnold LM, Lu Y, Crofford LJ, Wohlreich M, Detke MJ, Iyengar S, et al. A double-blind, multicenter trial comparing duloxetine with placebo in the treatment of fibromyalgia patients with or without major depressive disorder. *Arthritis Rheum* 2004;50:2974-84.
  34. Aaron LA, Bradley LA, Alarcon GS, Alexander RW, Triana-Alexander M, Martin MY, et al. Psychiatric diagnoses in patients with fibromyalgia are related to health care-seeking behavior rather than to illness. *Arthritis Rheum* 1996;39:436-45.
  35. Carville SF, Arendt-Nielsen S, Bliddal H, Blotman F, Branco JC, Buskila D, et al. EULAR evidence-based recommendations for the management of fibromyalgia syndrome. *Ann Rheum Dis* 2008;67:536-41.
  36. Hauser W, Eich W, Herrmann M, Nutzinger DO, Schiltenwolf M, Henningsen P. Fibromyalgia syndrome: classification, diagnosis, and treatment. *Dtsch Arztebl Int* 2009;106:383-91.
  37. Wolfe F, Hassett A, Walitt B, Michaud K. Mortality in fibromyalgia: an 8,186 patient study over 35 years. *Arthritis Care Res* 2011;63:94-101.