Development of Responder Definitions for Fibromyalgia Clinical Trials

Lesley M. Arnold,[1] David A. Williams,[2] James I. Hudson,[3] Susan A. Martin,[4] Daniel J. Clauw,[2] Leslie J. Crofford,[5] Fujun Wang,[6] Birol Emir,[7] Chinglin Lai,[8] Rong Zablocki,[9] and Philip J. Mease[10]

Objective. To develop responder definitions for fibromyalgia (FM) clinical trials using key symptom and function domains.

Methods. Twenty-four candidate responder definitions were developed by expert consensus and were evaluated in 12 randomized, placebo-controlled trials of 4 medications for the treatment of FM. For each definition, the treatment effects of the medication compared with placebo were analyzed using Cochran-Mantel-Haenszel tests or chi-square tests. A meta-analysis of the pooled results for the 4 medications established risk ratios to determine the definitions that best favored medication over placebo.

Results. Two definitions performed best in the analyses. Both definitions included ≥30% reduction in pain and ≥10% improvement in physical function. The definitions differed in that one (≥30% improvement in FM [FM30] short version) included ≥30% improvement in sleep or fatigue, and the other (FM30 long version) required ≥30% improvement in 2 of the following symptoms: sleep, fatigue, depression, anxiety, or cognition. In the analysis of both versions, the response rate was ≥15% for each medication and was significantly greater compared with placebo. The risk ratio favoring drug over placebo in the pooled analysis for FM30 version 3 (short version) was 1.50 (95% confidence interval [95% CI] 1.24–1.82; $P \leq 0.0001$); the risk ratio for FM30 version 6 (long version) was 1.60 (95% CI 1.31–1.96; $P \leq 0.00001$).

Conclusion. Among the 24 responder definitions tested, 2 were identified as most sensitive in identifying response to treatment. The identification of responder

definitions for FM clinical trials that include assessments of key symptom and function domains may improve the sensitivity of clinical trials to identify meaningful improvements, leading to improved management of FM.

Fibromyalgia (FM) is defined by the American College of Rheumatology (ACR) as widespread pain (duration of ≥3 months) in combination with tenderness at 11 or more of 18 specific tender point sites (1). Clinically, and in clinical trials of therapy for FM, responder indices of successful outcomes should more broadly address the associated symptoms of fatigue and cognitive dysfunction, sleep and mood disturbances, and lowered functional status that influence a patient's perception of whether his or her FM has "improved" (2–4).

Currently, 3 medications are approved by the US Food and Drug Administration for the management of FM, including the $\alpha_2\delta$ ligand pregabalin and the serotonin and norepinephrine reuptake inhibitors duloxetine and milnacipran (5–7). Recent trials have shown the efficacy of other medications for the treatment of FM, and continued development of new treatments is likely (8). Evaluating the comparative efficacy of interventions for FM is difficult, however, because no common definition of response in FM exists. At present, the inclusion of assessment domains is inconsistent, and there is wide variation in the use of instruments indexing those domains.

An empirically derived responder definition would facilitate the aggregation of multiple clinically important outcomes into a single metric that could then serve as a primary outcome in clinical trials. Clinical decision-making could also be based on this common metric rather than requiring clinicians to make inferences about a given patient from group means in reference samples across multiple symptom domains. The responder approach also helps identify whether improvement in key outcomes occurs within the same person; such identification is a clinical necessity when evaluating the treatment response in a condition with the complexities of FM (9). A responder definition also facilitates the prediction of individual responses to treatments, which is an important aid to long-term treatment planning and management of this chronic condition.

Historically, many symptoms have been thought to be associated with FM. Because an assessment of all symptoms in each patient is not feasible, consensus was required to identify the key domains that needed to be assessed to determine clinically meaningful improvement. Much of the work in this area has been organized by the Outcome Measures in Rheumatology (OMERACT) FM working group (3). OMERACT is an international organization representing a partnership between academic clinicians, industry, and government agencies sharing a common interest in promoting the development of the best possible outcome measures for clinical trials affecting rheumatologic conditions (10). Delphi exercises involving both clinicians and patients were conducted by the OMERACT FM working group members (11,12). From both the consensus process by clinicians and patients and the confirmation process by analysis of clinical trials (13,14), a core domain set for FM assessment in clinical trials and practice was established and ratified by OMERACT. The core domains included pain, tenderness, fatigue, patient global assessment of change, multidimensional function, and sleep disturbance. Other important domains included depression, cognitive dysfunction, stiffness, and anxiety (3).

Most clinical trials in FM have used a global measure of improvement known as the Patient Global Impression of Change (PGIC) scale. The PGIC scale can be used as an index of improvement against which each of the domain-specific measures can be mapped, thus revealing the association between overall improvement and specific symptom domains. Data derived from FM clinical trials of pregabalin, gabapentin, duloxetine, and milnacipran enabled the OMERACT FM working group to evaluate the association of change within specific symptom domains with patients' overall impression of improvement. Results from these analyses further supported some of the consensus-derived OMERACT domains as being the most salient domains associated with patients' perceptions of overall improvement. In the pregabalin trials, pain, fatigue, sleep, and work and physical function were the most pronounced independent predictors of improvement in PGIC (15). Similarly, pain, fatigue, sleep, and physical and social function were predictors of improvement in an FM trial of gabapentin (16). In an assessment of data from clinical trials of duloxetine, independent predictor variables of end point PGIC included assessments for pain, physical function, fatigue, anxiety, social function, and tender point thresholds (17). Among responders to milnacipran, improvement in pain, fatigue, sleep, physical function, and cognitive symptoms were associated with better PGIC ratings (18). Taken together, the most consistent ranking of domains associated with PGIC included pain, fatigue, physical function, and sleep.

The present study was designed to develop and test candidate definitions of improvement in FM using preexisting FM clinical trial databases, by adopting some

of the approaches used to develop the definition of improvement in rheumatoid arthritis (i.e., the ACR 20%, 50%, 70% improvement criteria) (19) and by applying the data-driven consensus methods advocated by OMERACT (10). The goal was to combine important symptom and function domains into a responder definition, using outcome measures that were most common across FM clinical trials and that were shown to be valid and sensitive to change (14,20).

MATERIALS AND METHODS

Guided by the approach used in the development of the ACR definition of improvement in RA (19), the OMERACT FM working group identified initial drafts of candidate responder definitions by consensus during the European League Against Rheumatism 2009 Congress and at the 2009 ACR annual meeting. Six versions of responder definitions were identified based on the previously reviewed core set of clinically relevant domains (Table 1). Following the approach used by the ACR to determine criteria for improvement in RA, each of the 6 draft versions contained 4 possible definitions depending on the percent improvement ($\geq$20%, $\geq$30%, $\geq$50%, and $\geq$70% [FM20, FM30, FM50, and FM70, respectively]) in outcome measures, for a total of 24 candidate responder definitions.

Two approaches were used for the selection of domains to be included in responder definitions. First, some of the candidate definitions included only the domains of pain, physical function, fatigue, and sleep, which were most consistently associated with PGIC in the studies reviewed above. The second approach included additional domains to reflect the heterogeneity of the FM population and the recognition that some treatments may affect other domains of importance such as depression, anxiety, and cognition. None of the candidate definitions included an assessment of tenderness, because tender points were not consistently evaluated in the clinical trials of FM and, as observed in trials to date, do not appear to be sensitive to change (21,22). Although cognitive dysfunction was also not consistently measured in the clinical trials, this domain was highly ranked in importance by patients, and recent trials have begun to explore assessment of this common symptom (23). Finally, stiffness was not included in the candidate definitions, because many patients report stiffness as part of their overall pain experience.

Given that all treatments are unlikely to improve all symptom domains, the definitions included a requirement to meet the response criteria for pain and physical function but allowed for flexibility in other symptom domains. In addition, results of previous clinical trials in FM suggest that although improvement in function is critical, in 3-month trials, the level of improvement in function may not be the same as that in other symptom domains. Thus, some of the definitions allowed for a level of improvement in physical function (i.e., $\geq$10%) that was less than the expected level of improvement in pain (3).

When selecting outcome measures for the candidate definitions, we were limited to those used in the clinical trials to which we had access. Because of a lack of consensus regarding key FM clinical domains and outcome measures before the trials were conducted, a variety of measures were included in the trials. To make the assessment of the candidate responder indices as consistent as possible, we included the Fibromyalgia Impact Questionnaire (FIQ) (24) or the Short Form 36 (SF-36) (25), which were the only measures that were used in all of the trials. Although the individual items on the FIQ were not originally intended to be used in isolation to measure specific symptoms, changes in the 10 subscale items of the FIQ have been reported in recent studies (24). Furthermore, the domains of sleep, fatigue, depression, and anxiety were not otherwise assessed in some of the trials used in the present analysis. For most of the definitions, the SF-36 physical function domain was selected for evaluation of physical function, which was the area of function that, among the various dimensions of function, appeared to have the greatest influence on patient-reported improvement in PGIC analyses (15–18).

For the assessment of pain, we included either a numeric rating scale (score range 0–10) or a visual analog scale (VAS; score range 0–100) of pain severity. Cognition was assessed in the trials of only one of the medications, using the attention subscale of the Multiple Ability Self-Report Questionnaire, which contains items related to problems with attention and concentration that are commonly reported by patients with FM (26).

We recognized that the domains of sleep and fatigue have multiple dimensions, which supports the use of multidimensional measures that capture the various types of sleep and fatigue difficulties experienced by patients with FM. Therefore, to determine whether substituting other measures for sleep and fatigue would yield consistent results, we tested alternative candidate responder definitions for versions 1 and 3 that used multidimensional measures for these domains. Possible alternative fatigue measures included the Global Fatigue Index from the Multidimensional Assessment of Fatigue (27) and the Multidimensional Fatigue Inventory (28). Another measure for sleep was the sleep disturbance subscale of the Medical Outcomes Study (29).

Table 1 presents the candidate responder definitions. Each of the 6 versions was tested using 4 different levels of improvement (total of 24 definitions). The percent improvement in versions 1, 2, and 5 was the same across all domains and ranged from $\geq$20% (FM20) to $\geq$70% (FM70). In versions 3, 4, and 6, the percent improvement in pain and associated symptoms was the same in each definition (FM20, FM30, FM50, and FM70), but the percent improvement in physical function was fixed at $\geq$10%.

By consensus of the OMERACT FM working group, the best definitions were determined by the following criteria: 1) versions in which the patient response rate was $\geq$15% across all of the drugs and was significantly greater compared with placebo, and 2) versions in which the risk ratios of all drugs had a lower limit of the 95% confidence interval (95% CI) of $\geq$1.00 in the meta-analysis.

The candidate definitions were tested using data from 12 placebo-controlled clinical trials of 4 medications for the treatment of FM. Based on agreement with the sponsors of these trials, identifiable trial information was not included in the analyses. Each definition was first evaluated using data

Table 1. Candidate fibromyalgia responder definitions*

	Pain	Physical function	Associated symptoms
Version 1	≥20% (or ≥30%, ≥50%, ≥70%) reduction (NRS or VAS)	≥20% (or ≥30%, ≥50%, ≥70%) improvement in SF-36 physical function	≥20% (or ≥30%, ≥50%, ≥70%) improvement in one of the following: fatigue (FIQ tiredness) or sleep (FIQ rested)
Version 2	≥20% (or ≥30%, ≥50%, ≥70%) reduction (FIQ pain)	≥20% (or ≥30%, ≥50%, ≥70%) improvement in FIQ physical function	≥20% (or ≥30%, ≥50%, ≥70%) improvement in one of the following: fatigue (FIQ tiredness) or sleep (FIQ rested)
Version 3	≥20% (or ≥30%, ≥50%, ≥70%) reduction (NRS or VAS)	≥10% improvement in SF-36 physical function	≥20% (or ≥30%, ≥50%, ≥70%) improvement in one of the following: fatigue (FIQ tiredness) or sleep (FIQ rested)
Version 4	≥20% (or ≥30%, ≥50%, ≥70%) reduction (FIQ pain)	≥10% improvement in FIQ physical function	≥20% (or ≥30%, ≥50%, ≥70%) improvement in one of the following: fatigue (FIQ tiredness) or sleep (FIQ rested)
Version 5	≥20% (or ≥30%, ≥50%, ≥70%) reduction (NRS or VAS)	≥20% (or ≥30%, ≥50%, ≥70%) improvement in SF-36 physical function	≥20% (or ≥30%, ≥50%, ≥70%) improvement in 2 of the following: fatigue (FIQ tiredness), sleep (FIQ rested), depression (FIQ depression), anxiety (FIQ anxiety), or cognition (MASQ)
Version 6	≥20% (or ≥30%, ≥50%, ≥70%) reduction (NRS or VAS)	≥10% improvement in SF-36 physical function	≥20% (or ≥30%, ≥50%, ≥70%) improvement in 2 of the following: fatigue (FIQ tiredness), sleep (FIQ rested), depression (FIQ depression), anxiety (FIQ anxiety), or cognition (MASQ)

* NRS = numeric rating scale; VAS = visual analog scale; SF-36 = Short Form 36 Health Survey; FIQ = Fibromyalgia Impact Questionnaire; MASQ = Multiple Ability Self-Report Questionnaire.

from the 4 medications in order to identify the responder definition that best differentiated treatments from placebo.

All data from the placebo-controlled FM trials were pooled for each of the 4 medications. Only data from the short-term placebo-controlled phase (~3-month duration) were included, and efficacious doses were combined in the analysis. Intent-to-treat analyses were performed, imputing missing values by using the last observation carried forward. Patients for whom outcome measure data were missing at baseline or thereafter and for whom the response criteria could not be defined were excluded from the analysis; however, those with missing outcome measure data were included if the response could be defined from other measures. Each candidate response definition was assessed using a Cochran-Mantel-Haenszel test or a chi-square test to compare the treatment effects of the medication versus placebo while controlling for study effect.

In addition to the individual analyses of the 4 medications, meta-analyses of the pooled results (summary statistics) of all of the studies of the 4 medications were performed to establish risk ratios to help identify the response definitions that best favored the drug over placebo and to establish

discriminant validity (sensitivity to change) of the definitions. A random-effects model was used to estimate the overall effect.

RESULTS

Table 2 summarizes the results of the analyses of the pooled FM clinical trials for the 4 medications. The definitions in which the response rate was ≥15% and significantly greater than placebo across all drugs were FM20 versions 1, 3, 5, and 6 and FM30 versions 3 and 6. Thus, versions 3 and 6 performed the best when considering both the FM20 and FM30 definitions. Because the FM50 and FM70 candidate definitions had relatively few responders across all of the medications, these definitions were not included in the subsequent meta-analysis.

Table 3 shows the results of the meta-analysis of the pooled data for each of the medications that established risk ratios to help determine the definitions that

Table 2. Response rates among pooled fibromyalgia trials of 4 medications, using candidate responder definitions*

| | Version | | | | | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
	Placebo	Drug	Placebo	Drug	Placebo	Drug	Placebo	Drug	Placebo	Drug	Placebo	Drug
Drug A												
FM20	16.6	25.3†	20.8	28.7†	20.3	30.6‡	23.6	31.6†	16.6	26.1‡	19.5	32.7‡
FM30	11.7	17.1	12.7	20.0†	16.8	24.2§	16.9	24.7†	11.1	17.5§	15.6	25.9‡
FM50	5.5	9.9§	6.9	11.9†	11.3	17.1§	10.7	17.9†	5.5	12.1‡	10.7	20.2‡
FM70	2.1	4.1	3.4	6.0	5.5	9.4§	5.2	9.3§	1.8	4.2	5.5	10.1§
Drug B												
FM20	11.8	17.4‡	24.4	31.0‡	16.4	22.8‡	27.6	33.6†	12.5	18.5‡	17.3	25.1‡
FM30	6.5	10.9‡	17.3	23.0‡	12.9	18.0‡	21.0	27.6‡	7.2	12.0‡	13.9	20.5‡
FM50	2.6	4.1§	9.0	13.3†	8.1	11.6†	13.2	17.8†	2.7	4.4§	8.3	13.2‡
FM70	0.8	1.2	3.3	6.8‡	3.7	4.5	6.6	10.9‡	0.9	1.2	3.7	5.4§
Drug C												
FM20	15.4	31.5‡	28.0	44.2‡	20.8	38.4‡	29.4	47.9‡	15.5	31.3‡	20.8	37.9‡
FM30	9.4	21.4‡	23.9	35.5‡	16.4	34.1‡	27.1	41.0‡	9.1	21.8‡	15.8	34.6‡
FM50	3.8	10.3‡	13.7	23.2‡	9.7	25.5‡	15.5	29.9‡	4.7	10.1†	10.7	25.2‡
FM70	2.2	3.8	6.7	12.2†	6.6	15.5‡	9.0	18.7‡	2.2	4.1	6.9	16.6‡
Drug D												
FM20	17.5	21.4§	20.8	23.8	19.5	24.4§	23.1	25.9	17.8	22.2§	19.9	25.2§
FM30	11.2	14.6§	15.2	17.8	14.5	18.5§	18.9	21.0	12.0	14.8	15.1	19.3§
FM50	4.8	7.2	7.7	10.0	8.3	10.5	11.2	14.2	5.3	7.2	8.6	10.7
FM70	1.3	2.0	3.5	3.9	2.8	3.9	5.4	6.2	1.3	2.4	2.9	4.4

* Values are the percentage of responders. FM20 = ≥20% improvement in fibromyalgia symptoms.
† $P \leq 0.01$ versus placebo.
‡ $P \leq 0.001$ versus placebo.
§ $P \leq 0.05$ versus placebo.

best favored drug over placebo. The versions in which the risk ratios of all drugs had a lower limit of the 95% CI of ≥1.00 were FM20 versions 1, 3, 5, and 6 and FM30 versions 1, 3, and 6. Table 4 summarizes the meta-analysis results for all of the medications combined. Versions 1, 3, 5, and 6 had the highest risk ratios among the definitions, and FM30 had consistently higher risk ratios than FM20 in each version. FM30 versions 1 and 5 had higher risk ratios with smaller heterogeneity (I^2) values compared with FM30 versions 3 and 6, respectively. However, the response rates (Table 2) were lower for FM30 versions 1 and 5, with some drugs having a response rate of <15%. The lower response rates for versions 1 and 5 may be related to the higher threshold for improvement in physical function required in these versions. Therefore, the meta-analysis results, combined with the response rates, suggested that versions 3 and 6 are best for defining the response to any

Table 3. Meta-analysis of pooled data from all of the medications for selected candidate responder definitions*

| | Version | | | | | |
	1	2	3	4	5	6
Drug A						
FM20	1.52 (1.20−1.93)	1.38 (1.12−1.70)	1.51 (1.22−1.85)	1.34 (1.11−1.62)	1.57 (1.24−1.98)	1.68 (1.36−2.06)
FM30	1.46 (1.09−1.95)	1.57 (1.20−2.07)	1.44 (1.13−1.82)	1.46 (1.15−1.84)	1.58 (1.17−2.13)	1.66 (1.31−2.12)
Drug B						
FM20	1.47 (1.20−1.81)	1.27 (1.11−1.45)	1.39 (1.17−1.65)	1.22 (1.07−1.38)	1.48 (1.22−1.80)	1.45 (1.23−1.71)
FM30	1.69 (1.28−2.23)	1.33 (1.13−1.57)	1.40 (1.15–1.70)	1.31 (1.13−1.52)	1.67 (1.28−2.17)	1.48 (1.23−1.78)
Drug C						
FM20	2.04 (1.54−2.71)	1.58 (1.31−1.91)	1.85 (1.46−2.34)	1.63 (1.36−1.95)	2.03 (1.53−2.69)	1.82 (1.44−2.31)
FM30	2.26 (1.56−3.28)	1.49 (1.20−1.84)	2.09 (1.59−2.73)	1.51 (1.24−1.84)	2.38 (1.64−3.47)	2.19 (1.66−2.89)
Drug D						
FM20	1.23 (1.00−1.51)	1.14 (0.95−1.38)	1.25 (1.03−1.51)	1.12 (0.94−1.34)	1.25 (1.02−1.52)	1.27 (1.05−1.53)
FM30	1.30 (1.00−1.70)	1.17 (0.93−1.47)	1.28 (1.02−1.60)	1.11 (0.91−1.36)	1.23 (0.96−1.59)	1.28 (1.02−1.60)

* Values are the risk ratio (95% confidence interval). FM20 = ≥20% improvement in fibromyalgia symptoms.

Table 4. Meta-analysis of pooled data for all of the medications combined for selected candidate responder definitions*

	RR (95% CI)	Heterogeneity, %	P†
Version 1			
FM20	1.52 (1.25–1.83)	64	<0.0001
FM30	1.60 (1.30–1.98)	51	<0.0001
Version 2			
FM20	1.33 (1.17–1.51)	51	<0.0001
FM30	1.37 (1.22–1.53)	13	<0.00001
Version 3			
FM20	1.47 (1.26–1.70)	56	<0.00001
FM30	1.50 (1.24–1.82)	63	<0.0001
Version 4			
FM20	1.31 (1.13–1.52)	69	0.0005
FM30	1.33 (1.18–1.52)	43	<0.00001
Version 5			
FM20	1.53 (1.27–1.83)	62	<0.00001
FM30	1.63 (1.28–2.08)	64	<0.0001
Version 6			
FM20	1.52 (1.31–1.77)	57	<0.00001
FM30	1.60 (1.31–1.96)	68	<0.00001

* RR = relative risk; 95% CI = 95% confidence interval; FM20 = ≥20% improvement in fibromyalgia symptoms.
† Obtained from a random-effects model using the Cochran-Mantel-Haenszel test.

of the drugs compared with placebo. For both versions 3 and 6, FM30 definitions had greater risk ratios compared with the FM20 definitions. Figure 1 presents the forest plots for FM30 versions 3 and 6.

We also analyzed FM30 version 3 using alternative sleep and fatigue measures. For 2 of the drugs, alternative multidimensional measures for both fatigue and sleep were used across most of the trials. In the meta-analyses of these trials that included alternative measures, the total risk ratio for the alternative FM30 was 1.37 (95% CI 1.15–1.64, $P = 0.0004$).

DISCUSSION

The development of candidate responder definitions for FM trials is the result of several years of preliminary work that included multiple steps. We reviewed previous criteria for a response to treatment in FM involving multiple symptom domains. Several early definitions have limited utility now, because they included various assessments of tender points, which have
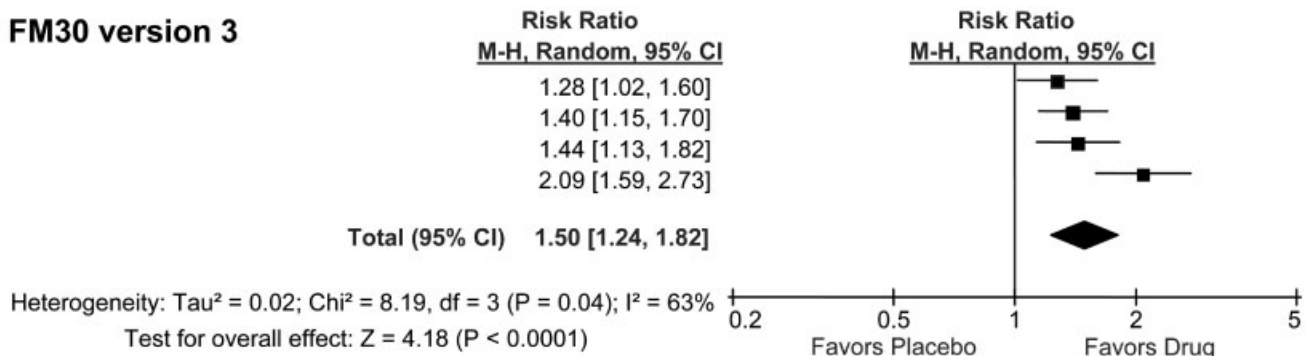
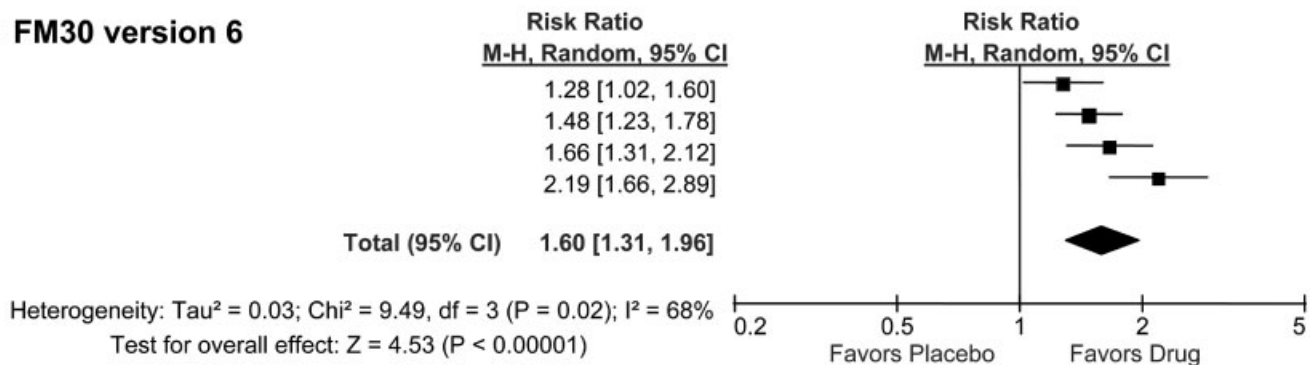**Figure 1.** Forest plots of **A,** FM30 (≥30% improvement in fibromyalgia symptoms) version 3 and **B,** FM30 version 6. Risk ratios and 95% confidence intervals (95% CIs) are presented, along with the results for heterogeneity. The overall effect was obtained using a random-effects model with Mantel-Haenszel (M-H) tests. Both versions showed that all drugs were statistically significantly superior compared with placebo.

not been consistently included as outcome measures in recent pivotal FM trials (30–33). In addition, some of the definitions did not include a change in pain, which is a cardinal feature of FM (32), or many of the other key symptom or function domains now recognized as clinically relevant (33).

More recently, composite responder definitions were used in FM trials of milnacipran and sodium oxybate. In the milnacipran trials, the composite responder definition for the treatment of FM consisted of 3 components: 1) ≥30% improvement from baseline in pain, 2) a rating of "very much improved" (score = 1) or "much improved" (score = 2) on the PGIC scale, and 3) ≥6-point improvement from baseline in physical function (SF-36 Physical Component Summary [PCS] score) (25). For the treatment of pain associated with FM, a 2-measure composite response definition included only the pain and PGIC components described above (34). The response definitions used in the milnacipran trials addressed some of the limitations of the previously proposed response criteria by including an assessment of function, but they did not include many of the other clinically relevant symptom domains identified by the OMERACT FM working group (3). In addition, the PCS score is a summary measure of multiple scales of the SF-36 designed to assess physical health, including body pain, and does not measure physical function alone. In the sodium oxybate trials, the composite responder definition was as follows: 1) ≥30% improvement in the VAS score for pain, 2) ≥30% improvement in the FIQ total score, and 3) a rating of "much better" or "very much better" on the PGIC scale (35). As in the case for the milnacipran studies, the responder definition used in the sodium oxybate studies did not contain many of the other key symptom domains as defined by the OMERACT FM working group. Although the FIQ total score reflects the impact of multiple symptoms (including pain) and function domains, it does not provide information about an individual's response to specific domains (24).

To address the limitations of previous FM responder definitions, we adopted some of the approaches used to develop the ACR definition of improvement in RA (19) and applied the data-driven consensus methods advocated by OMERACT (10). Because we recognized that FM is associated with multiple symptom and function domains, we identified key domains of FM through patient focus groups and patient and clinician Delphi exercises (11,12). After a review of the results of these patient and clinician studies and a confirmation process by analysis of clinical trials (13,14), a list of core domains

was selected by consensus as part of an OMERACT module on FM (3). We then conducted analyses of existing FM clinical trial databases to determine which of the domains drove patients' perception of improvement (15–18). We evaluated the performance characteristics of outcome measures for the core domains to assess their face, construct, content, and criterion validity as well as sensitivity to change (14,20). Finally, in the present study, we combined core symptom and function domains to develop candidate responder definitions, using outcome measures that were most commonly used across FM clinical trials and were found to be valid and sensitive to change. We tested candidate responder definitions using existing FM clinical trial databases.

The candidate definitions included a combination of key symptom and function domains. We observed that the responder definitions that best favored drug over placebo included improvement in pain and physical function as well as improvement in either sleep or fatigue (FM30 version 3 [short version]). Along with pain, sleep disturbance and fatigue have been consistently ranked by patients and clinicians as being among the most common and troublesome symptoms of FM (3). The other responder definition that performed well in the analysis (FM30 version 6 [long version]) included additional symptom domains of depression, anxiety, and cognitive dysfunction to reflect the heterogeneity of the FM population and the recognition that some treatments may affect these other domains of importance. A responder definition that includes improvement in specific key symptom domains in addition to pain evaluates the broader impact of FM on patients and addresses the limitations of other composite responder definitions for FM trials that focused only on the symptom of pain.

The best responder definitions (FM30 versions 3 [short version] and 6 [long version]) included at least 30% improvement in the symptom domains. This is consistent with other studies that showed that at least 30% improvement in pain represents a clinically important improvement in chronic pain disorders (36,37), including FM (38). There are few published data on clinically important changes in other FM symptom domains such as sleep and fatigue, and to our knowledge, there are no published data on clinically important changes in multiple symptom domains that are part of a composite responder definition. However, in support of requiring the same level of improvement in other symptom domains (e.g., ≥30%) as is required for the pain domain, a recent pooled analysis of duloxetine FM clinical trials showed that patients who reported feeling

much or very much better had similar levels of improvement in pain, sleep, and fatigue (39).

For the assessment of symptoms of sleep and fatigue, we included definitions that used 10 FIQ single items for morning tiredness (a measure of refreshing sleep) and fatigue. We also tested multidimensional measures for sleep and fatigue and observed that the results were consistent with findings that used single FIQ items. Future studies should evaluate the responder definitions using new outcome measures such as those being developed within the National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS) initiative, a collaborative effort to develop patient-reported outcomes for a wide variety of chronic diseases and conditions (40).

FM affects multiple dimensions of function, but analyses of previous clinical FM trials suggested that physical function had the greatest influence on patient-reported improvement (15–18). Therefore, when function was assessed by the responder definitions, only an assessment of physical function was used. Both FM30 versions 3 and 6 used the SF-36 physical function scale to assess function. Versions that used the alternative measure of physical function, the FIQ function component, did not perform as well as the versions using the SF-36. This may be related to the observation that the FIQ function component is oriented toward individuals with severe FM and high levels of disability, resulting in a potential floor effect for those with milder disease (41). Recently, a revised FIQ in which the function questions were modified was shown to be better correlated with the SF-36 physical function scale and may be considered for inclusion as an outcome measure in future studies of the responder definitions, along with other measures of function developed through PROMIS (42).

Several limitations of this study should be considered. First, as noted above, the analyses were based on completed FM clinical trials that were conducted before consensus was reached on key symptom and function domains in FM. Therefore, we were limited in the choice of outcome measures when developing the responder definitions, and some of the important domains (e.g., cognition) were not consistently evaluated across the trials. In addition the trials used variable measures to assess symptoms such as sleep disturbance and fatigue. However, our analyses suggest that different measures of a clinical domain can be substituted in the definitions, as is done in the ACR responder criteria for RA.

Second, all of the trials used the FIQ and the SF-36, which allowed for some consistency in evaluating the domains. As noted, the SF-36 appeared to be a better assessment of physical function compared with the FIQ; however, the individual FIQ items for sleep and fatigue performed similarly to more multidimensional scales, suggesting that simple numeric scales may be sufficient for the evaluation of some domains.

Third, there was little difference between several of the definitions in terms of their ability to detect a response to medication. Therefore, the definitions chosen should be regarded as a useful first step in the development of response criteria based on this method.

Fourth, although we obtained consensus from OMERACT about the appropriateness of the proposed responder definitions (4), we have not yet obtained patient feedback on the responder definitions, which is planned for the future.

Finally, in addition to responder definitions, there is a need for an assessment of worsening and flare as well as a measure for monitoring disease activity over time, as has been recommended by OMERACT (4). These additional studies will be described in a future report.

The FM responder definitions that were identified as the most sensitive in identifying response to treatment in analyses of existing clinical trials of 4 medications in FM included FM30 version 3 (short version) and version 6 (long version). These definitions share common features in that they require ≥30% reduction in pain and ≥10% improvement in physical function. FM30 version 3 also requires ≥30% improvement in sleep or fatigue, while FM30 version 6 requires ≥30% improvement in 2 of the following symptoms: sleep, fatigue, depression, anxiety, or cognition. These responder definitions can be used to improve the assessment of a patient's response to treatment. Future studies should explore the use of these responder definitions in FM clinical trials.

AUTHOR CONTRIBUTIONS

ROLE OF THE STUDY SPONSORS

statistical support. Publication of this article was not contingent upon approval by Eli Lilly, Pfizer, Jazz Pharmaceuticals, Cypress Bioscience, or Forest Laboratories.

ADDITIONAL DISCLOSURES

Author Wang is an employee of Eli Lilly. Author Emir is an employee of Pfizer. Author Lai is an employee of Jazz Pharmaceuticals. Author Zablocki is an employee of Cypress Bioscience.

REFERENCES

1. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the Multicenter Criteria Committee. Arthritis Rheum 1990;33:160–72.
2. Williams DA, Clauw DJ. Understanding fibromyalgia: lessons from the broader pain research community. J Pain 2009;10:777–91.
3. Mease P, Arnold LM, Choy EH, Clauw DJ, Crofford LJ, Glass JM, et al. Fibromyalgia syndrome at OMERACT 9: domain construct. J Rheumatol 2009;36:2318–29.
4. Mease PJ, Clauw DJ, Christensen R, Crofford LJ, Gendreau RM, Martin SA, et al. Toward development of a fibromyalgia responder index and disease activity score: OMERACT module update. J Rheumatol 2011;38:1487–95.
5. Lyrica (pregabalin) [package insert]. New York: Pfizer; 2010.
6. Cymbalta (duloxetine) [package insert]. Indianapolis: Eli Lilly; 2010.
7. Savella (milnacipran) [package insert]. St. Louis (MO): Forest Pharmaceuticals; 2010.
8. Arnold LM. Biology and therapy of fibromyalgia: new therapies in fibromyalgia. Arthritis Res Ther 2006;8:212.
9. Witter J, Simon LS, Dianne R. Are means meaningless? The application of individual responder analysis to analgesic drug development. APS Bull 2003;13:4–7.
10. Boers M, Brooks P, Simon LS, Strand V, Tugwell P. OMERACT: an international initiative to improve outcome measurement in rheumatology. Clin Exp Rheumatol 2005;23:S10–3.
11. Arnold LM, Crofford LJ, Mease PJ, Burgess SM, Palmer SC, Abetz L, et al. Patients' perspective on the impact of fibromyalgia. Patient Educ Couns 2008;73:114–20.
12. Mease PJ, Arnold LM, Crofford LJ, Williams DA, Russell IJ, Humphrey L, et al. Identifying the clinical domains of fibromyalgia: contributions from clinician and patient Delphi exercises. Arthritis Rheum 2008;59:952–60.
13. Mease P, Arnold LM, Bennett R, Boonen A, Buskila D, Carville S, et al. Fibromyalgia syndrome: OMERACT 8 workshop. J Rheumatol 2007;34:1415–25.
14. Choy EH, Arnold LM, Clauw DJ, Crofford LJ, Glass JM, Simon LS, et al. Content and criterion validity of the preliminary core dataset for clinical trials in fibromyalgia syndrome. J Rheumatol 2009;36:2330–4.
15. Arnold LM, Zlateva G, Sadosky A, Emir B, Whalen E. Correlations between fibromyalgia symptom and function domains and patient global impression of change: a pooled analysis of three randomized, placebo-controlled trials of pregabalin. Pain Med 2011;12:260–7.
16. Arnold L, Martin S, Welge J. Correlations between fibromyalgia symptom and function domains and the patient global impression of improvement: an analysis of the randomized, placebo-controlled trial of gabapentin [abstract]. Ann Rheum Dis 2010;69 Suppl III:iii445.
17. Hudson JI, Arnold LM, Bradley LA, Choy EH, Mease PJ, Wang F, et al. What makes fibromyalgia patients feel better? Correlations between patient global impression of improvement and changes in clinical symptoms and function: a pooled analysis of four randomized, placebo-controlled trials of duloxetine. J Rheumatol 2009;36:2517–22.
18. Geisser ME, Clauw DJ, Strand V, Gendreau RM, Palmer R, Williams DA. Contributions of change in clinical status parameters to Patient Global Impression of Change (PGIC) scores among persons with fibromyalgia treated with milnacipran. Pain 2010;149:373–8.
19. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
20. Carville SF, Choy EH. Systematic review of discriminating power of outcome measures used in clinical trials of fibromyalgia. J Rheumatol 2008;35:2094–105.
21. Arnold LM, Keck PE Jr, Welge JA. Antidepressant treatment of fibromyalgia: a meta-analysis and review. Psychosomatics 2000;41:104–13.
22. O'Malley PG, Balden E, Tomkins G, Santoro J, Kroenke K, Jackson JL. Treatment of fibromyalgia with antidepressants: a meta-analysis. J Gen Intern Med 2000;15:659–66.
23. Mease PJ, Arnold L, Wang F, Ahl J, Mohs R, Gaynor P, et al. The effect of duloxetine on cognition in patients with fibromyalgia [abstract]. Arthritis Rheum 2010;62 Suppl:S42.
24. Bennett R. The Fibromyalgia Impact Questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. Clin Exp Rheumatol 2005;23:S154–62.
25. 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.
26. Seidenberg M, Haltiner A, Taylor MA, Hermann BB, Wyler A. Development and validation of a Multiple Ability Self-Report Questionnaire. J Clin Exp Neuropsychol 1994;16:93–104.
27. Belza B, Henke C, Yelin E, Epstein WV, Gilless CL. Correlations of fatigue in older adults with rheumatoid arthritis. Nurs Res 1993;42:93–9.
28. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315–25.
29. Hays RD, Stewart AL. Sleep measures. In: Stewart AL, Ware JE, editors. Measuring functioning and well-being. Durham (NC): Duke University Press; 1992. p. 235–9.
30. Simms RW, Felson DT, Goldenberg DL. Criteria for response to treatment in fibromyalgia [abstract]. Arthritis Rheum 1988;31 Suppl:S100.
31. Carette S, Bell MJ, Reynolds WJ, Haraoui B, McCain GA, Bykerk VP, et al. Comparison of amitriptyline, cyclobenzaprine, and placebo in the treatment of fibromyalgia: a randomized, double-blind clinical trial. Arthritis Rheum 1994;37:32–40.
32. Simms RW, Felson DT, Goldenberg DL. Development of preliminary criteria for response to treatment in fibromyalgia syndrome. J Rheumatol 1991;18:1558–63.
33. Dunkl PR, Taylor AG, McConnell GG, Alfano AP, Conaway MR. Responsiveness of fibromyalgia clinical trial outcome measures. J Rheumatol 2000;27:2683–91.
34. Arnold LM, Gendreau RM, Palmer RH, Gendreau JF, Wang Y. Efficacy and safety of milnacipran 100 mg/day in patients with fibromyalgia: results of a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2010;62:2745–56.
35. Russell IJ, Perkins AT, Michalek JE, and the Oxybate SXB-26 Fibromyalgia Syndrome Study Group. Sodium oxybate relieves pain and improves function in fibromyalgia syndrome: a randomized, double-blind, placebo-controlled, multicenter clinical trial. Arthritis Rheum 2009;60:299–309.
36. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–58.

37. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.
38. Mease PJ, Spaeth M, Clauw DJ, Arnold LM, Bradley LA, Russell IJ, et al. Estimation of minimum clinically important difference of pain in fibromyalgia. Arthritis Care Res 2011;63:821–6.
39. Ang D, Oakes TM, Jarvis J, Wang F. Correlation of overall Patient Impression of Improvement with individual items in the Fibromyalgia Impact Questionnaire [abstract]. J Pain 2010;11 Suppl:S38.
40. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al, on behalf of the PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PRO-MIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010;63: 1179–94.
41. Wolfe F, Hawley DJ, Goldenberg DL, Russell IJ, Buskila D, Neumann L. The assessment of functional impairment in fibromyalgia (FM): Rasch analyses of 5 functional scales and the development of the FM health assessment questionnaire. J Rheumatol 2000;27:1989–99.
42. Bennett RM, Friend R, Jones KD, Ward R, Han BK, Ross RL. The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties [published erratum appears in Arthritis Res Ther 2009;11:415]. Arthritis Res Ther 2009;11: R120.