

Unhelpful Criteria Sets For “Diagnosis” and “Assessment of Severity” of Fibromyalgia



In this issue of *The Journal* Dr. Wolfe and colleagues introduce 2 new and different protocols (there have been many prior versions) to identify what they call “fibromyalgians”^{1,2}. In what ways are the new proposals unhelpful? Dilution (the inclusion of milder cases changes prognosis); inconsistency (multiple diagnostic strategies identify different patients); loss of specificity [later criteria sets lose information and fail to discriminate symptoms of rheumatoid arthritis (RA), fibromyalgia (FM), systemic lupus erythematosus (SLE), and osteoarthritis (OA)]; and loss of the ability to recognize FM concurrent with other diseases. Statistical approaches are used that do not make most efficient use of the information available. More fundamental is the continuing distraction from the obligation to accurately identify the origins of pain and choose appropriate treatments³.

The odyssey was not without purpose or logic. FM has been described as a spectrum disorder. But there are 2 quite separate spectra. One is a continuous spectrum expressing severity, with FM separated from lesser chronic pain by arbitrary cutpoints. The other is better described as a palette of characteristic symptoms, found to be associated with FM however defined. Many of these statistically significant associations were found in the 1990 Classification Criteria study⁴. Sleep disturbance, fatigue, morning stiffness > 15 minutes, paresthesias, anxiety, headache, prior depression, irritable bowel syndrome, sicca symptoms, urinary urgency, dysmenorrhea history, and Raynaud’s phenomenon are all listed in Table 6 of the 1990 paper. Because symptoms other than widespread pain were not embedded in the Criteria, the statistical strength of the association of these and other symptoms associated with FM could be tested independent of the 1990 Criteria. Other links were added later, including cognitive problems (“Fibrofog”)⁵, “dizziness,” jaw pain^{6,7}, and low abdominal pain. If this symptom profile is related strongly to tender point counts, perhaps the profile could be inverted to infer tender point counts, and allow a valid diagnosis of FM without tender points. A very obvious precedent

is David Sackett’s “number-needed-to-treat” (NNT). The NNT is simply the reciprocal of the absolute risk difference and is widely used and cited. But the answer has a hyperbolic rather than linear or Gaussian distribution (not specified in the resulting publications). When there is no treatment effect, the absolute risk reduction is zero and the NNT is infinite⁸.

The evaluation of any one “new” criteria set has to begin with the studies that led to the selection of the 1990 Criteria for the Classification of Fibromyalgia³. The data from that study, and in many later studies, were presented as a receiver-operating characteristic (ROC) curve. These (and related curves, e.g., the Gini coefficient) present 2 problems; the plots consume most or all of the degrees of freedom (are “penalized”), leaving few or none for error assessments. Further, they assume without examination that distances between multiple curves are distributed normally with constant standard deviations. The curves are better examined after transformation to permit analysis as hyperbolae or power curves⁴. Explanation of all this introduces may not be easy reading, but is necessary to the examination of the statistical treatment of later suggested criteria sets.

In Figure 1, the curve on the left is a ROC curve, plotted from the point-count data in Figure 2 of the 1990 American College of Rheumatology (ACR) report⁴. The points are fitted by a spline curve that gives great accuracy but leaves no degrees of freedom for statistical testing. The graph on the right has the same data and axes, but is rescaled so that maximum rather than minimum values are bottom left. This transformation allows parametric analysis as a power or hyperbolic curve (see Figure 2).

In Figure 2, in the left graph, the data in Figure 1 are analyzed as a power curve. The model uses only 2 degrees of freedom, and 22 are available to plot the 95% confidence limits. R-squared values for fit (original scale) are 0.971. The log-log plot on the right closely approximates a straight line, with narrow and almost uniform confidence limits.

Defining and validating a new approach to the diagnosis

*See FM criteria and measuring severity, page 1113
and Do we need core sets of FM domains? page 1104*

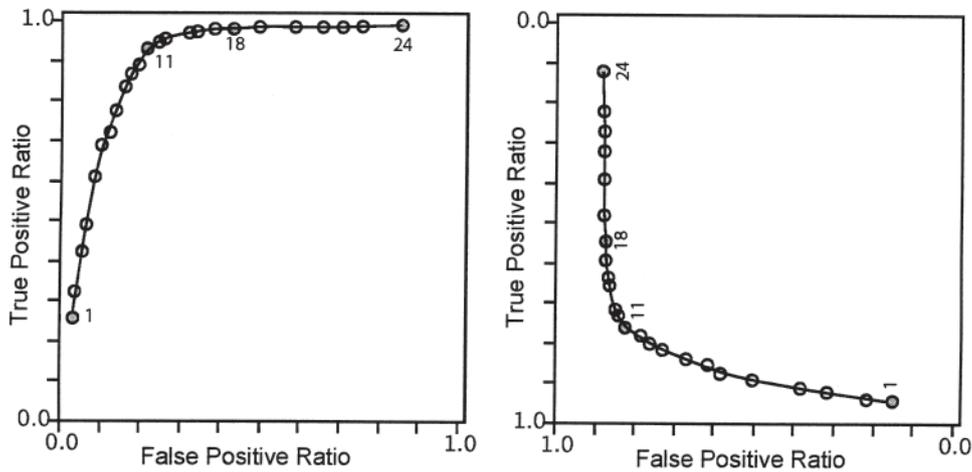


Figure 1. ROC curve plotted from point-count data in the 1990 ACR classification of FM report⁴ (left panel) and same data and axes rescaled (right panel) to allow parametric analysis (see text).

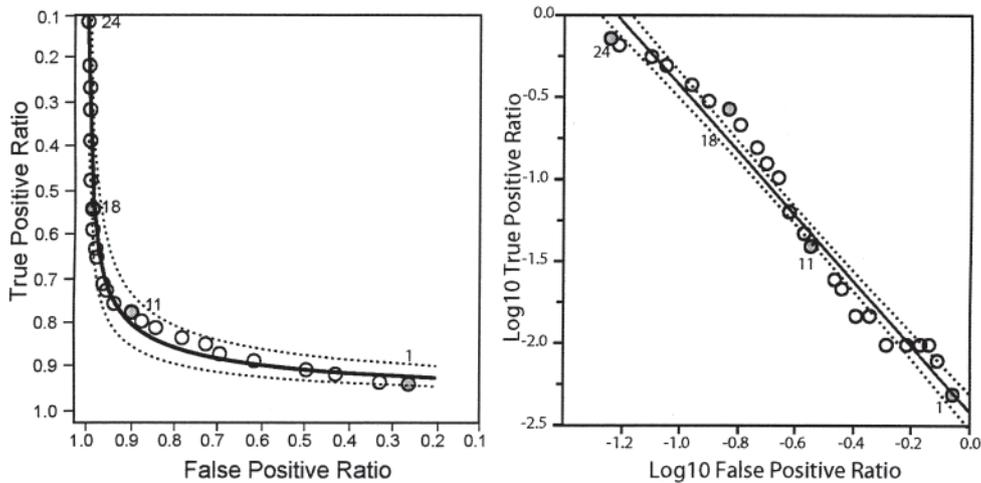


Figure 2. Data from 1990 ACR classification of FM report⁴ analyzed as a power curve (left panel) and log-log plot (right panel) (see text).

of FM is not to be undertaken lightly. In the general population, there will be about 50 people without FM for every one that qualifies. The symptoms listed above are not uncommon; some 30% of adults in truly random samples (your control group) will have musculoskeletal pain, and at least half of these will have neck and/or back pain. Sample sizes will need to be very large. Each recruit will have to fill out a lengthy questionnaire, which then must be e-mailed and scored. You may then learn that the “ACR” protocol used is now out of date, and that the current one is under revision.

The concept was dramatically introduced by Wolfe in 2003⁹. In building his database, the National Data Bank for Rheumatic Diseases, he had to accept input from physicians who could not or would not do tender point counts. But it seems that the data also lacked a specific definition of the characteristic palette of symptoms listed above. What they

had was a “combination of fatigue, the regional pain score, a count of somatic symptoms, and a count of lifetime comorbid illnesses”: not the same thing at all. Nor was this “composite” specific to FM; patients with RA had the same profile, to a lesser extent. Diagnostic specificity was lost. The same measure was used for “diagnosis” as for assessment of severity.

We then had 3 different criteria sets: the 1990 criteria (accepted by Wolfe and many colleagues as the “de facto standard”¹⁰), the “expert” clinician’s diagnosis, and the proposed Survey Criteria. Did these criteria identify the same patients? In a study published in 2006¹¹, only half (60 of 120) of patients identified as FM by any of the 3 approaches met the 1990 Criteria. Of the 60 not meeting the ACR criteria, the clinician and the Survey Criteria agreed on only 26, and disagreed on 34. Diagnosis based on alternative

approaches seriously inflated (diluted) the number identified with patients not recognized by the 1990 standard, with different and/or milder illness, and therefore different prognosis and treatment needs. If there had been an interval between the assessments, the diagnosis of the clinician would have been accepted as “true fibromyalgia,” and the negative judgment of the Survey Criteria would justify a new category, true “prior fibromyalgia,” with no “false-positives.”

In 2010 a further attempt was published¹². “Physicians enrolled 258 (263 in Table 1) valid patients whose clinical diagnosis was FM and 256 (251 in Table 1) who were control subjects.” “It was not a requirement of (prior) diagnosis to have satisfied the ACR classification criteria.” Of the 196 patients with “current fibromyalgia,” 184 met 1990 criteria as applied by their physician, with a mean tender point count of 15.9. A group of 67 failed to meet those criteria (the mean point count was 8). These “false-positives” were labeled “prior fibromyalgia” without information about prior point counts. They did have high “somatic symptom scores” in the 3 months prior to assessment.

What “somatic symptoms”? Not the restricted palette of characteristic symptoms validated in the 1990 study and a few later studies, but any of “blurred vision or problems focusing; dry eyes; ringing in ears; hearing difficulties; mouth sores; dry mouth; loss of or change in taste; headache; dizziness; fever; chest pain; shortness of breath; wheezing (asthma); loss of appetite; nausea; heartburn; indigestion or belching; pain or discomfort in the upper abdomen (stomach); liver problems; pain or cramps in the lower abdomen (colon); diarrhea (frequent, explosive watery bowel movements, severe); constipation; black or tarry stools (not from iron); vomiting; joint pain; joint swelling; low back pain; muscle pain; neck pain; weakness of muscles; tiredness (fatigue); depression; insomnia; nervousness (anxiety); seizures or convulsions; trouble thinking or remembering; easy bruising; hives or welts; itching; rash; loss of hair; red, white, and blue skin color changes in fingers on exposure to cold or with emotional upset; sun sensitivity (unusual skin reaction, not sunburn); yellow skin or eyes (jaundice); fluid-filled blisters; numbness/tingling/burning; swelling of the hands, legs, feet, or ankles (not due to arthritis); irritable bowel syndrome; faintness; frequent urination; painful urination; pain, fullness, or discomfort in the bladder region; sensitivity to bright lights, loud noises, or odors; fatigue severe enough to limit daily activity; tender lymph nodes; or frequent sore throats”¹¹. Some fit into the *Portrait of Dorian Gray* as painted by Ivan Albright. Mercifully, this protocol has not persisted.

A confusion develops between the use of scores to measure severity, and the same scores to establish diagnosis. The accompanying graphs share with ROC curves a lack of statistical structure, with no degrees of freedom for error estimates.

The bold further revision in this issue of *The Journal*¹ proposes that “a ‘slight’ modification to the ACR2010 (criteria) will allow their use in epidemiologic and clinical studies without the requirement for an examiner”. Physicians are entirely unnecessary; the somatic symptoms scored have been reduced to 3, and a new “FM Symptom scale (FS)” was created “by summing the Widespread Pain Index (WPI) and the modified Symptom Severity scale”. “This scale has also been called the fibromyalgiance scale” (a phrase repeated 3 times in the text). Again, the same tool is used to make the “diagnosis,” and to measure severity. Function is not assessed. For the latter purpose, it is inferior to the revised Fibromyalgia Impact Scale¹³ and other measures.

“The criteria are simple to use and administer, but they are not to be used for self-diagnosis” or for medicolegal purposes. (But of course they will be, if endorsed by the American College of Rheumatology.) Given the dilution of ACR90-positive patients with patients with milder ACR90-negative symptoms, they should be used in therapeutic and other research studies only with respect for their limitations, evolving structure, and alternatives.

The FM Symptom Scale predominately measures pain, and when used in patients with RA, SLE, OA, and FM, is diagnostically without any features (other than pain) that differentiate the “fibromyalgiance” from pain related to the primary disease.

One comment is worth noting; “The exact wording of the depression question could be a matter of concern. In the context that we used the word depression, it meant depressive symptoms, feelings of depression, or depressed mood. It was not to indicate a medical diagnosis of depression.” This is consistent with the data collected in the evolution of the scoring of the Medical Outcome Study Short-Form 36 in truly random samples of general populations¹⁴, in which pain was “normally” associated with anxiety and depression, and scoring weights were adjusted accordingly.

The second article² in this issue concludes: “The main determinants of global severity and health related quality of life in FM are pain, function, and fatigue. But these variables are also the main determinants in RA and other rheumatic diseases. The content and impact of FM, whether measured by discrete variables or a fibromyalgiance scale, seems to be independent of diagnosis. These data argue for a common set of variables rather than disease-specific variables.” This seems to imply that the variables discussed lack specificity. But the text uses the word “fibromyalgiance” 18 times!

HUGH A. SMYTHE, MD, FRCP,
Associate Editor,
The Journal of Rheumatology,
Toronto, Canada

Address correspondence to Dr. H.A. Smythe, 2 Heathbridge Park,
Toronto, Ontario M4G 2Y6, Canada. E-mail: hasmythe@rogers.com

REFERENCES

1. Wolfe F, Clauw DJ, Fitzcharles M-A, Goldenberg DL, Häuser H, 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;38:1113-22.
2. Wolfe F, Hassett AL, Katz RS, Michaud K, Walitt B. Do we need core sets of fibromyalgia domains? The assessment of fibromyalgia (and other rheumatic disorders) in clinical practice. *J Rheumatol* 2011;38:1104-12.
3. Smythe HA. Explaining medically unexplained symptoms: widespread pain. *J Rheumatol* 2009;36:679-83.
4. 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.
5. Leavitt F, Katz RS, Mills M, Heard AR. Cognitive and dissociative manifestations in fibromyalgia. *J Clin Rheumatol* 2002;8:77-84.
6. Wolfe F, Katz RS, Michaud K. Jaw pain: its prevalence and meaning in patients with rheumatoid arthritis, osteoarthritis, and fibromyalgia. *J Rheumatol* 2005;32:2421-8.
7. Smythe HA. Temporomandibular joint disorder and other medically unexplained symptoms in rheumatoid arthritis, osteoarthritis, and fibromyalgia. *J Rheumatol* 2005;32:2288-90.
8. Smythe HA. Percent change was and is an embarrassing mistake. *J Rheumatol* 2010;37:2448-51.
9. Wolfe F. Stop using the American College of Rheumatology criteria in the clinic. *J Rheumatol* 2003;30:1671-2.
10. Wolfe F. The fibromyalgia syndrome: a consensus report on fibromyalgia and disability. *J Rheumatol* 1996;23:534-9.
11. Katz RS, Wolfe F, Michaud K. Fibromyalgia diagnosis: A comparison of clinical, survey, and American College of Rheumatology criteria. *Arthritis Rheum* 2006;54:160-76.
12. 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.
13. Bennett RM, Friend R, Jones KD, Ward R, Han BK, Ross RL. The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties. *Arthritis Res Ther* 2009;11:R120.
14. Ware JE, Kosinski M. SF-36 Physical and Mental Health Summary Scales; a manual for users of version 1. 2nd ed. Lincoln, RI: QualityMetric Incorporated; 2001.

J Rheumatol 2011;38:975-8; doi:10.3899/jrheum.110142