An Artificial Neural Network Approach for Predicting Functional Outcome in Fibromyalgia Syndrome after Multidisciplinary Pain Program

Monika Salgueiro, PhD,¹ Xabier Basogain, Dr-Ing,² Antonio Collado, MD, PhD,³ Xavier Torres, PhD,³ Juan Bilbao, Dr-Ing,¹ Francisco Doñate, MD, PhD,¹ Luciano Aguilera, MD, PhD,⁴ and Jon Jatsu Azkue, MD, PhD⁵

Departments of ¹Neurosciences and ²Preventive Medicine and Public Health, School of Medicine and Dentistry, University of the Basque Country, Leioa;
³Department of Systems Engineering and Automatics, Faculty of Engineering, University of the Basque Country, Bilbao;
⁴Fibromyalgia Unit;
⁵Clinical Psychology Service, Hospital Clínic i Provincial de Barcelona, Barcelona;
⁶Department of Surgery, Radiology and Physical Medicine, University of the Basque Country;
⁷Anesthesiology and Reanimation Service, Hospital de Basurto, Bilbao, Spain

Reprint requests to: Jon Jatsu Azkue, MD, PhD, Department of Neurosciences, School of Medicine and Dentistry, University of the Basque Country, Barrio Sarriena s/n, 48940 Leioa, Spain. Tel: +34-94601-8379; Fax: +34-94601-5055; E-mail: jonjatsu.azkue@ehu.es.

The authors declare no conflict of interest (AKA dual commitments, competing interests, and competing loyalties).

Abstract

Objective. The objective of this study was to evaluate the ability of artificial neural networks (ANNs) to predict, on the basis of clinical variables, the response of persons with fibromyalgia syndrome (FMS) to a standard, 4-week interdisciplinary pain program.

Design. The design of this study is retrospective longitudinal.

Setting. Fibromyalgia outpatient clinic in a tertiary-care general hospital.

Subjects. The subjects of this study include outpatients with FMS.

Intervention. Multidisciplinary pain program including pain pharmacotherapy, cognitive-behavioral therapy, physical therapy, and occupational therapy.

Outcome Measures. Reliable change (RC) of scores on the Stanford Health Assessment Questionnaire (HAQ), and accuracy of ANNs in predicting RC at discharge or at 6-month follow-up as compared to Logistic Regression.

Results. ANN-based models using the sensory-discriminative and affective-motivational subscales of the McGill Pain Questionnaire, the HAQ disability index, and anxiety subscale of Hospital Anxiety and Depression Scale at baseline as input variables correctly classified 81.81% of responders at discharge and 83.33% of responders at 6-month follow-up, as well as 100% of nonresponders at either evaluation time-point. Logistic regression analysis, which was used for comparison, could predict treatment outcome with accuracies of 86.11% and 61.11% at discharge and follow-up, respectively, based on baseline scores on the HAQ and the mental summary component of the Medical Outcomes Study—Short Form 36.

Conclusions. Properly trained ANNs can be a useful tool for optimal treatment selection at an early stage after diagnosis, thus contributing to minimize the lag until symptom amelioration and improving tertiary prevention in patients with FMS.

Key Words. Fibromyalgia; Pain; Treatment Outcome

Introduction

Fibromyalgia syndrome (FMS) is a highly prevalent disorder characterized by chronic widespread pain, disturbed sleep, cognitive symptoms, fatigue, and a number of somatic symptoms [1]. As primary and secondary prevention are not viable due to the lack of biological markers and limited knowledge of the underlying pathogenic mechanisms, rapid access to treatment and functional rehabilitation is essential to minimize negative impact of
established disease. However, reaching an optimal treatment in FMS is largely empirical and usually involves substantial trial and error and patient monitoring for several months each time a new treatment is initiated, and this delay in achieving symptomatic improvement is often the cause of prolonged frustration and suffering [2–4]. In addition, persons with FMS may differ considerably in their psychosocial profiles [5,6] and their perception of pain [7,8], and there are sizable individual differences in responses to treatment [9,10].

In the current scenario, customizing therapeutic intervention to patients’ clinical profiles holds promise to improve treatment outcome. Progress toward this goal will require the development of clinical prediction rules, i.e., empirical statements that are formulated to improve the efficiency and accuracy of clinical judgments [11,12], which will ultimately assist the clinician in adopting an optimal therapeutic approach on the basis of patient’s characteristics and clinical settings. The first step in this process is the identification of predictors of treatment outcome from observational studies [11].

In this work, we have explored the ability of artificial neural networks (ANNs) to predict patient response to a standard, 4-week interdisciplinary pain program on the basis of a variety of clinical baseline characteristics as input data. ANNs are a widely available modeling technique that compares favorably with multiple linear general models for classifying dependent variables in a number of complex data models and proves useful for classification in complex biological systems when prediction depends on interactions of many variables [13–15]. An ANN can be trained to recognize specific patterns of input data and to provide a classification label as an output. We show that ANN-based models can attain excellent predictive performance that can be potentially exploited for clinical decision making.

Methods

Setting and Baseline Data Collection

A total of 72 outpatients with FMS (69 women and 3 men) aged 24–59 (mean age 41.5 years) attending the Fibromyalgia Unit at the Hospital Clinic i Provincial in Barcelona participated in the study (Table 1). The study was approved by the institutional ethical review boards of all three participating institutions, and all participant subjects provided their written, informed consent. Sociodemographic data were collected first, and questionnaires were then administered in the same session for baseline clinical assessment, including the Fibromyalgia Impact Questionnaire (FIQ), Medical Outcomes Study—Short Form 36 (SF-36), the Stanford Health Assessment Questionnaire (HAQ), the Hospital Anxiety and Depression Scale (HADS), and the McGill Pain Questionnaire (MPQ) (Table 2). All questionnaires were first administered at baseline, and then at discharge and at follow-up 5 months later.

### Table 1 Sociodemographic characteristics of the study group

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>69</td>
<td>95.83</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>4.16</td>
</tr>
<tr>
<td>Age group (years)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21–30</td>
<td>7</td>
<td>9.72</td>
</tr>
<tr>
<td>31–40</td>
<td>25</td>
<td>34.72</td>
</tr>
<tr>
<td>41–50</td>
<td>35</td>
<td>48.61</td>
</tr>
<tr>
<td>51–60</td>
<td>5</td>
<td>6.94</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>21</td>
<td>29.16</td>
</tr>
<tr>
<td>High school</td>
<td>31</td>
<td>43.05</td>
</tr>
<tr>
<td>University</td>
<td>20</td>
<td>27.77</td>
</tr>
<tr>
<td>Activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>38</td>
<td>52.77</td>
</tr>
<tr>
<td>Inactive</td>
<td>34</td>
<td>47.22</td>
</tr>
</tbody>
</table>

* Mean age 41.5 years; range 24–59 years.

The FIQ is a 20-term, self-administered questionnaire that assesses both psychological and physical symptoms of FMS and allows for a broad quantification of interference of the condition on daily living tasks and quality of life [16]. It is quickly administered and easily evaluates a broad range of clinical characteristics associated to FMS. The total FIQ score can adopt values ranging from 0 to 10, where higher scores denote greater severity or impact. We used the Spanish language version validated by Monterde and colleagues [17].

The Spanish version of the SF-36 [18] was administered to assess overall health status and health-related quality of life (HRQOL). The physical component summary (PCS)
and the mental component summary (MCS) were calculated from scorings on appropriate items of the test. PCS and MCS scorings can range from 0 to 100, where higher values relate to better quality of life.

The HAQ is a self-assessment scale that has been developed as a tool for detecting states of depression and anxiety in the setting of a hospital outpatient clinic. It includes anxiety and depressive subscales. We used the HAQ version validated for Spanish language [19]. The HAQ (cf. below) is a commonly used tool in rheumatology quantifying difficulty in performing activities of daily living and measuring functional disability in patient groups [20,21]. We used the validated Spanish translation of the HAQ [22].

The MPQ is a widely employed verbal pain assessment tool. It comprises 20 subscales or sets of verbal descriptors designed to evaluate the sensory, affective, evaluative, and miscellaneous dimensions of pain [23]. The MPQ also includes a visual analog scale (VAS) as current pain self-assessment tool. A Pain Rating Index (PRI) was calculated from weighted MPQ item scores for each of the pain dimensions addressed by the questionnaire, i.e., sensory-discriminative, affective-motivational, evaluative, and miscellaneous dimensions of pain. Only the sensory and affective indices were included for study. We used the validated Spanish translation of the MPQ [24].

Interdisciplinary Treatment

FMS outpatients included in this study underwent a multidisciplinary pain program in the Fibromyalgia Unit of the Hospital Clinic i Provincial de Barcelona under supervision of the institutional Rheumatology, Anesthesiology, Rehabilitation, Psychiatry and Clinical Psychology Services of the Hospital Clinic [25]. The final protocol was further reviewed by the Scientific Committee of the Spanish Foundation of Fibromyalgia and Chronic Fatigue Syndrome patients, along with members of the Spanish and Catalan Scientific Societies of Rheumatology, Rehabilitation and Physical Medicine, Internal Medicine, Clinical Psychology, Family and Community Medicine, and the Spanish Society of Pain. Criteria for inclusion were 1) age between 18 and 65 years, 2) no major psychopathology (severe suicide risk, schizophrenia or other psychotic disorders, or severe personality disorder), 3) no drug dependence/abuse, and 4) not being in the course of labor litigation or application for disability benefits. Briefly, the multidisciplinary intervention encompassed pain pharmacotherapy (based on pregabalin, serotonin-norepinephrine reuptake inhibitors, as well as nonsteroidal anti-inflammatory drugs or opioids on an as-needed basis), cognitive-behavioral therapy (motivational interviewing, electromyographic biofeedback-assisted relaxation training, pacing techniques, distraction and visualization techniques, cognitive restructuring, and assertive training), physical therapy, and occupational therapy (training in daily activities, simulation of work tasks, modification of the risk factors at work, and ergonomic redesign of the workplace as needed). As many of our patients entered treatment with a number of misconceptions regarding pain, an important component of this multidisciplinary program was to provide accurate, reliable information via medical, psychological, occupational, and ergonomic seminars. The treatment was delivered in 12 sessions over a 4-week period on a 3-days-a-week basis (5 hours a day).

Functional Outcome Measure

In this study, the HAQ was used to evaluate functional status and treatment outcome. The HAQ measures difficulty in performing activities of daily living and is commonly used in rheumatology to quantify functional disability in patient groups [20,21,26]. This instrument contains 20 questions, classified into eight domains (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities), regarding functional limitations due to a disease condition. From these, the HAQ disability index (DI) ranging from 0 to 3 can be computed as a sensitive measure of disability, where scores of 0–1 are generally considered to represent mild to moderate difficulty, 1–2 moderate to severe disability, and 2–3 severe to very severe disability. An advantage of this tool is that normative data for the general population and patient populations are available. The estimated population mean HAQ-DI is 0.25 (95% confidence interval [95% CI] 0.22–0.28) [27], whereas average scores that have been reported in osteoarthritis and rheumatoid arthritis patients are 0.8 and 1.2, respectively [21].

We used the reliable change (RC) statistic to evaluate if the change in scores as a result of treatment was greater than would be expected from random variation alone. The RC index (RCI) determines whether the patient’s status has changed reliably by calculating if the magnitude of the difference between the initial score and the score after treatment is greater than a certain expected level corrected for the standard error of the measurement. As described by Jacobson and Truax [28], the RCI is defined as the difference between pretest and posttest scores, divided by the standard error of the measurement. The latter is measured as $SD\sqrt{2/(1-r)}$, where SD1 is the standard deviation of the sample at baseline and r is the test–retest reliability of the Spanish-validated version of the instrument ($r=0.89$ [22]). A spreadsheet (LibreOffice Calc v. 3.6.2.2, http://www.libreoffice.org/about-us/credits/) was used here to calculate RCI. In this study, patients with a positive RC, i.e., those whose improvement was greater than the measurement error for the course of treatment to be considered successful (RCI greater than 1.96) were be labeled as responders (P < 0.05), whereas patients either showing a negative RC (RCI < −1.96) or falling within the band of no RC (−1.96 to 1.96) were labeled as nonresponders.

ANN Analysis

We sought to reach parsimonious ANN-based models exhibiting acceptable predicting capabilities. Therefore,
we tentatively built and trained a number of feed-forward neural networks of the multilayer perceptron class, from which three final models were selected on the basis of their performances on training and validation phases. Architecturally, the networks were composed of an input layer consisting of as many computational processing elements—the so-called neurons—as selected variables from the baseline patient assessment (Table 2), a hidden neurons layer, and an output layer providing the result of the network’s prediction in terms of a binomial variable (Figure 1). In the basic multilayer perceptron architecture, each input to a neuron is multiplied by a weight, and inputs to a given neuron plus a bias term are then all summed. Weights and biases are assigned random values prior to learning, and the learning phase consists of iteratively providing the input layer with selected clinical data along with the desired output, so the network internally adjusts weights associated with each neuron according to the differences between obtained and desired output values. Here, patient scores from the various baseline assessment tools were normalized and fed into input layers, and heuristic methods were used for ANN training. In all cases, ANNs were trained for 50,000 cycles by using a back-propagation training algorithm, the delta rule as learning rule, and a sigmoid function for transforming the total net input to a neuron to obtain its final output. In the three final ANN models, the root mean square errors at the output layer were close to 0.2. Out of the total of 72 subjects included in this study, data from two thirds of participating subjects (48 individuals) were randomly included in the learning phase of ANN development (training data). The predictive performance of developed ANN-based models was then evaluated in terms of positive predictive value (precision rate), negative predictive value, and prediction accuracy by presenting the networks with unused data from the remaining 24 individuals (test data).

Logistic Regression Analysis

Logistic regression (LR) analyses were carried out for comparison to the performance of ANN-based models. All subjects in the sample were included. Stepwise binomial regression analyses were performed by combining forward selection and backward elimination approaches in parallel with ANN analysis and by using the same input data as for ANN training (Table 2), with positive RC at discharge or at 6-month follow-up as dependent variables. Variables were retained only if they contributed to a statistically significant model at $P < 0.05$. Positive predictive value (precision rate), negative predictive value, and prediction accuracy of the so-obtained regression models were calculated on the basis of predicted and actual outcome measures.

**Figure 1** Schematic representation of general architecture and input/output configuration of artificial neural networks (ANNs) built and trained in the present study. The network was composed of a four-neuron input layer receiving patient data, a hidden layer comprising five neurons, and an output layer providing the output result and containing one or two neurons. Model 3 is shown here, in which the output was provided by two neurons. In all three models trained, normalized scorings on the sensory-discriminative and affective-motivations subscales of the McGill Pain Questionnaire (MPQ), the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS), and the HAQ disability index (DI) were fed to neurons of the input layer of the network. During the training phase, an additional input channel (not depicted) introduced the desired output result to the network, i.e., whether or not the patient should be considered as responder.
Results

Characteristics of Responders and Nonresponders

Clinical variables used for patient evaluation and for building models to predict response to multidisciplinary treatment are reported in Table 2. Table 3 shows scorings of the entire patient group at baseline, along with measures of outcome segregated by responder and nonresponder categories both at discharge and at 6-month follow-up.

When comparing the changes in clinical evaluation variables after treatment relative to baseline, statistically significant improvements were found in the responder group at discharge evaluation in several clinical measures (paired-samples t-test), including pain (VAS), sensory dimension of the MPQ, the anxiety subscale of the HADS, FIQ scorings, the PCS of the SF-36, and predictably the HAQ-DI as basic criterion of patient classification (Table 3).

In addition, significant changes were noted in the responder group at 6-month follow-up relative to baseline regarding pain (VAS), impact of the disease in terms of FIQ, as well as the PCS and again the HAQ-DI (Table 3). In patients labeled as nonresponders, only the HAQ-DI was found to be slightly but significantly ($P < 0.05$) lowered at discharge relative to baseline (Table 3).

When comparing the clinical status of responders to that of nonresponders, significant differences (independent-samples t-test) were found at discharge between the two groups as to average intensity of pain (VAS) and scorings on the sensory-discriminative dimension of pain (MPQ), on the FIQ, on the anxiety subscale of the HADS, as well as on the HAQ-DI. We failed to detect statistically significant differences between responders and nonresponders at the 6-month follow-up.

ANN Analysis

Tentative ANNs were developed at the first instance by using a variety of network configurations and combinations of clinical variables used for baseline patient evaluation as input data. From these, three final models were selected on the basis of their performances on training and validation phases. In all three, the hidden layer was composed of five neurons and clinical variables fed into the input layer included scorings in the sensory-discriminative and affective-motivational subscales of the MPQ, the HAQ-DI, and the anxiety subscale of HADS (Figure 1). Models 1 and 2 were trained to predict whether or not a person with FMS about to enter the multidisciplinary intervention would exhibit positive RC at discharge and at 6-month follow-up, respectively. Model 1 correctly classified 81.81% of responders and 100% of nonresponders in the test data set, with an overall accuracy of 91.66% (Table 4). Model 2 correctly classified 83.33% of responders and 100% of nonresponders in the test data set, with an overall accuracy of 91.66% (Table 4). The

Table 3  Patient outcome among responders and nonresponders

<table>
<thead>
<tr>
<th>Clinical Variable</th>
<th>Baseline</th>
<th>Discharge</th>
<th>6-Month Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain (VAS)</td>
<td>6.94 ± 1.31</td>
<td>4.82 ± 1.86***</td>
<td>6.50 ± 1.83**</td>
</tr>
<tr>
<td>MPQ Sensory Pain Rating Index</td>
<td>25.15 ± 4.95</td>
<td>20.98 ± 5.89***</td>
<td>24.98 ± 5.42**</td>
</tr>
<tr>
<td>MPQ Affective Pain Rating Index</td>
<td>4.75 ± 2.19</td>
<td>3.32 ± 2.35</td>
<td>4.65 ± 2.31</td>
</tr>
<tr>
<td>Fibromyalgia Impact Questionnaire</td>
<td>7.01 ± 1.14</td>
<td>5.58 ± 1.58***</td>
<td>6.95 ± 1.60**</td>
</tr>
<tr>
<td>HADS anxiety subscale</td>
<td>12.43 ± 4.12</td>
<td>9.89 ± 3.61****</td>
<td>12.38 ± 4.25*</td>
</tr>
<tr>
<td>SF-36 mental component summary</td>
<td>46.25 ± 26.20</td>
<td>48.72 ± 23.38</td>
<td>45.29 ± 27.47</td>
</tr>
<tr>
<td>Health Assessment Questionnaire</td>
<td>1.73 ± 0.64</td>
<td>1.20 ± 0.36***</td>
<td>1.49 ± 0.50*,***</td>
</tr>
</tbody>
</table>

Clinical characteristics of the whole study group at baseline are presented, as measured by scorings on a battery of clinical evaluation instruments, as well as scoring segregated by responder and nonresponder categories at both discharge and 6-month follow-up. Data are given as mean ± standard deviation. Statistical significance of differences between responders and nonresponders at the two evaluation time-points (independent-samples t-test) are marked by one and two asterisks (* equals $P < 0.05$ and ** equals $P < 0.01$). Significant changes at discharge or at follow-up in either responders or nonresponders relative to status prior to treatment (paired-samples t-test) are marked by three and four asterisks (*** and **** denote $P$ values of $<0.05$ or $<0.01$, respectively).

HADS = Hospital Anxiety and Depression Scale; MPQ = McGill Pain Questionnaire; SF-36 = Medical Outcomes Study—Short Form 36; VAS = visual analog scale.
areas under the respective receiver operating characteristic (ROC) curves were 0.91 and 0.94, respectively (Figure 2A).

Model 3 was trained to predict the prospective pattern of response over the two evaluation time-points, i.e., whether the subject would be classifiable as responder or nonresponder at discharge and then at 6-month follow-up (Figure 1). The so-constructed model correctly classified 20 out of 24 cases (83.33%) in the test data set. Details on correctly classified cases across the four possible response patterns are reported in Table 5.

### Logistic Regression Analysis

For comparison purposes, all clinical variables used as input data for ANN analysis, i.e., all those used for patient baseline evaluation, were considered as potential explanatory variables in regression analyses. After using stepwise forward selection, backward elimination, and personal discretion, the best LR model for patient response at discharge included the HAQ scores and MCS from the SF-36 questionnaire at baseline evaluation as significant predictors ($B = 4.11$ and $0.11$; Wald = 11.30 and 9.75 with a 95% CI for HAQ and MCS, respectively; Table 6). The overall model was significant at $P < 0.01$ and correctly classified 75% of responders and 91.66% of nonresponders with an overall accuracy of 86.11% (Table 6). No other attempted sets of explanatory variables yielded a statistically significant model. For classifying patients as responders or nonresponders at 6-month follow-up, only HAQ scores at baseline evaluation showed explanatory value ($B = 1.89$ and Wald = 7.74 at 95% CI; Table 6). The model was significant at $P < 0.01$ and correctly classified 34.61% of responders and 76.08% of nonresponders, with an overall accuracy of 61.11%. Areas under the corresponding ROC curves for the two LR models were 0.907 and 0.70, respectively (Figure 2).

### Discussion

This is the first report investigating a potential role for ANNs in predicting the response to a standard multidisciplinary treatment in patients with FMS. The data suggest that the learning and adaptive capabilities of ANNs can be of use to classify persons with FMS as prospective responders or nonresponders on the basis of baseline clinical data.

#### Response to Multidisciplinary Intervention

Training an ANN with classificatory purposes involves presenting the network with the desired output, along with the input data, in such a manner that its internal neuronal connections are progressively adjusted to associate input profile with the correct output result [13]. In this phase, a critical step is that patients presented to the ANN as responders had been correctly labeled as such because the clinical improvement of patients identified by trained ANNs as prospective responders will be expected to be significant enough to justify having been selected for intervention. In this study, we addressed this issue by labeling study subjects as responders on the basis of their changes in functional clinical status after therapeutic intervention. In this regard, multimodal treatment is deemed as the currently most effective treatment option in FMS [2,29,30], and this is important in such a disease as FMS where despite the need for effective treatment options both for medical and economic reasons, no more than approximately one third of patients may derive some benefit from any treatment in place [31]. To assess the impact of treatment upon the patient, we relied on the HAQ to measure HRQOL longitudinally. The HAQ is, to the best of current knowledge, among the most sensitive patient-oriented outcome assessment instruments to detect clinical change in rheumatologic conditions [32–34], including FMS [35–37]. Moreover, we used the RC statistic as a principled criterion for identifying positive,
clinically meaningful changes in longitudinal course of clinical status. This posed more stringent conditions for labeling an individual as a responder. Because of this, patients labeled here as nonresponders also included those undergoing minor but significant improvement that did, however, not attained the RCI level. Such minor positive responses probably contributed to the slight but significant average HAQ-DI improvement noted in the nonresponder group at discharge (Table 3).

Comparison of clinical profiles between patients labeled here as responders and nonresponders showed consistent differences across sensory, psychological, and functional parameters that could be measured with different assessment tools. Thus, patients in the responder group exhibited improvement relative to prior to treatment in terms of average scores on the FIQ, pain intensity, sensory PRI, the anxiety subscale of the HADS, and the PCS from the SF-36 (Table 3), whereas patients in the nonresponder group experienced no improvement in any of these variables. Consistently, responders scored statistically significantly better than nonresponders on most of these scales, at least at discharge (Table 3). We thus believe that the observed differences in patient outcome between responders and nonresponders are confirmatory of the adequacy of criteria for group segregation. Consequently, new patients with FMS that will be classified as prospective responders by the ANN-based models built from these data can be expected to be those that will actually experience reliable, clinically relevant improvement.

Contribution of ANN-Based Models to Prediction of Treatment

The challenge of predicting treatment outcome is one of clinical relevance in FMS. In the management of persons with this condition, there is typically a prolonged lag between the initiation of treatment and the symptomatic response. Current clinical practice is to monitor the patient

Table 5  Summary of performance of an artificial neural network (ANN)-based model in predicting four patterns of reliable change (RC) in functional status over the two evaluation time-points considered in this study

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Frequency in Test Data Set</th>
<th>Classified Correctly (Model 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1</td>
<td>R both at discharge and follow-up</td>
<td>6/24</td>
</tr>
<tr>
<td>Pattern 2</td>
<td>R at discharge and NR at follow-up</td>
<td>4/24</td>
</tr>
<tr>
<td>Pattern 3</td>
<td>NR at discharge and R at follow-up</td>
<td>1/24</td>
</tr>
<tr>
<td>Pattern 4</td>
<td>NR both at discharge and follow-up</td>
<td>13/24</td>
</tr>
<tr>
<td>Total</td>
<td>20/24</td>
<td></td>
</tr>
</tbody>
</table>

R = responder; NR = nonresponder.

Data used for evaluation of model performance are only test data, i.e., new data unused for network training. The network is capable of correctly predicting the pattern of response in 83.33% of newly presented patients.
with FMS for symptomatic improvement for up to 8–12 weeks after treatment onset [2,3], a process that is commonly repeated on a trial-and-error basis until the best available treatment is reached. In addition, the uncertainty of benefit and the problem of side effects also cause many patients to prematurely discontinue a medication before experiencing benefit [38–41]. This poses a major limitation to tertiary prevention, which otherwise seeks to preclude further deterioration or to reduce complications after the disease has declared itself. Prediction of treatment outcome may thus be expected to contribute to tertiary prevention by helping prioritize patients appropriately so that a specific therapy is administered to those that are most likely to benefit from it. Here, patients showing worse functional status in terms of baseline HAQ-DI tended to gain more from multidisciplinary treatment. This is in line with previous findings in patients with rheumatologic conditions. For example, higher scores on pain intensity and depression at baseline in patients with FMS have been found to be prognostic for improvement as a result of multidisciplinary intervention [42]. In addition, worse clinical status in terms of scores of pain intensity and depression has been found to be prognostic for more improvement after a multidisciplinary rehabilitation program in patients with chronic low back pain [43]. Although not directly measured here, this may be because reducing the amount of fear-avoidance beliefs, an important factor worsening physical functioning of patients with pain [43–45], is reduced during multidisciplinary treatments that include cognitive-behavioral therapy as in the present study [46]. Thus, in prioritizing patients with FMS when prescribing a multimodal intervention, higher functional disability may reasonably be seen as a potentially useful broad criterion to select those patients that are likely to benefit the most from treatment. More specifically, the ANN-based models built and trained here to identify prospective responders to multidisciplinary treatment showed an accuracy above 90% in classifying newly presented patients and compared favorably to LR. The subset of patients (48 out of 72) included in the training phase of the ANNs was selected randomly from the general sample and thus differences in sample size might have only minor influence on these results. A limitation of using ANNs is their relative inability to explicitly identify possible causal relationships, as analysis of coefficients associated to hidden layer neurons usually reveals little information as to the relative weight of individual input variables on outcomes. It is, however, important to stress that the aim of the present study was to demonstrate the potential usefulness of ANN-based models rather than to identify predicting factors. Although LR analysis remains the gold standard of statistical predicting tools in clinical settings, achieving most stable, meaningful results when large data samples are available, ANN-based models here exhibited an excellent predictive performance in a relatively small cohort of patients with FMS. This can be accounted for by the ability of ANNs to implicitly detect and model arbitrarily complex nonlinear relationships between variables, a common scenario within the medical domain and clinical diagnosis [13–15]. Indeed, ANN-based models here were able to capture and map several clinical variables in the input space, such as the anxiety subscale of the HADS and the sensory-discriminative and affective-motivational subscales of the MPQ, which failed to be identified as explanatory variables by the LR analysis. This is of particular importance when dealing with patients with FMS, which are characterized by considerable heterogeneity in clinical presentations, frequent concurrence of overlapping, comorbid conditions [31], and variable time-course [47,48]. In addition, it is probably due to the ability of ANNs to identify complex patterns, regardless of the complexity of the underlying logic, that treatment outcome at the 6-month follow-up was predicted better than by LR. Reasonably, therefore, the so-trained ANN-based models may be able to effectively identify as prospective responders those patients that will actually benefit from multimodal therapeutic intervention, provided that appropriate clinical data are collected as input data to the ANN during the training phase. Importantly, this triage can be done at an early stage after diagnosis or shortly after admission, thus minimizing the lag until symptomatic improvement. In addition, reducing uncertainty on patient outcome should be expected to contribute to improving adherence to treatment.

Finally, we show that the learning and adaptive capabilities of ANNs may be of use in predicting patterns of patient outcome or clinical evolution over time, which in the

### Table 6  Stepwise logistic regression coefficients of predictive models at discharge and at 6-month follow-up

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>Significance</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discharge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Assessment Questionnaire</td>
<td>4.11</td>
<td>1.22</td>
<td>11.30</td>
<td>&lt;0.01</td>
<td>61.18</td>
</tr>
<tr>
<td>SF-36 Mental Component Summary</td>
<td>0.09</td>
<td>0.02</td>
<td>9.75</td>
<td>&lt;0.01</td>
<td>1.09</td>
</tr>
<tr>
<td><strong>Six-month follow-up</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Assessment Questionnaire</td>
<td>1.89</td>
<td>0.68</td>
<td>7.74</td>
<td>&lt;0.01</td>
<td>6.63</td>
</tr>
</tbody>
</table>

SE = standard error.

HAQ scoring at baseline evaluation is a major predictor of reliable change (RC) after multidisciplinary treatment at both evaluation time-points. Yet to a far lesser extent, the mental component summary (MCS) from the Medical Outcomes Study—Short Form 36 (SF-36) also contributes to model predicting treatment outcome at discharge.
The present case included combined responses at two different evaluation time-points after treatment, provided that ANNs are trained with appropriate longitudinal data.

**Toward Treatment Individualization**

There is tremendous variability within patients with FMS, and an explicit diagnosis of FMS using ACR criteria may not produce a homogeneous patient sample that can be effectively treated with one approach to intervention [8,49–51]. In this scenario, specific therapeutic strategies could be made more effective if targeted at identified subgroups of patients early in the treatment process [50]. Clinical data analyzed here were collected from outpatients attending a hospital-based facility, and therefore the predicting capabilities of our present ANN-based models are circumscribed to these particular settings and cannot be generalized to patients with FMS collectively. However, the potentialities of ANNs to identify prospective responders on the basis of clinical data at baseline may also as effectively be exploited to model patient outcome in different specific clinical or demographic settings and with respect to any particular treatment. A so-constructed multiscale predictive framework may thus further contribute to therapeutic decision making and can contribute to and to progress toward treatment individualization.

ANNs are an affordable, little time demanding technology that requires less formal training to develop than statistical models and involves only modest computational effort. In summary, the results of the present study suggest that ANN-based models may be a useful adjunct to classic, LR models to therapeutic decision making in FMS, as well as a valuable tool to progress toward treatment individualization.

**Acknowledgments**

The project was supported by VI Premio Fibromialgia 2010 by Fundación FF y Ciencia SER, as well as by the Government of the Basque Country under program SAIOTEK (SA-2010/00110) and Ayudas a Grupos de Investigación del Sistema Universitario Vasco. Monika Salgueiro was supported by the Government of the Basque Country.

**References**


18 Alonso J, Prieto L, Anto JM. [The Spanish version of
the SF-36 Health Survey (the SF-36 health question
naire): An instrument for measuring clinical results].

of the hospital anxiety and depression scale (HADS) in
a Spanish population. Gen Hosp Psychiatry 2003;

20 Bruce B, Fries JF. The Health Assessment
8.

21 Bruce B, Fries JF. The Stanford Health Assessment
Questionnaire: Dimensions and practical applications.

22 Esteve-Vives J, Battie-Gualda E, Reig A. Spanish
version of the Health Assessment Questionnaire: Reliabil
ity, validity and transcultural equivalency. Grupo para
la Adaptacion del HAQ a la Poblacion Española.

23 Melzack R. The McGill Pain Questionnaire: Major
99.

24 Lázaro C, Bosch F, Torrubia R, Baños J-E. The develop-
ment of a Spanish Questionnaire for assessing pain:
Preliminary data concerning reliability and validity. Eur

25 Collado A, Torres X, Arias A, et al. [Efficiency of multi-
disciplinary treatment of chronic pain with locomotor

26 Fries JF, Spitz P, Kraines RG, Holman HR. Measure-
ment of patient outcome in arthritis. Arthritis Rheum

27 Krishnan E, Sokka T, Hakkinen A, Hubert H, Han-
nonen P. Normative values for the Health Assessment
Questionnaire disability index: Benchmarking disability
in the general population. Arthritis Rheum 2004;
50:953–60.

28 Jacobson NS, Truax P. Clinical significance: A statis-
tical approach to defining meaningful change in psy-
59:12–9.

29 Hauser W, Bernardy K, Arnold B, Offenbacher M,
Schiltenwolf M. Efficacy of multicomponent treatment
in fibromyalgia syndrome: A meta-analysis of random-
ized controlled clinical trials. Arthritis Rheum 2009;

30 Rossy LA, Buckelew SP, Dorr N, et al. A meta-analysis
of fibromyalgia treatment interventions. Ann Behav

31 Turk DC. Fibromyalgia: A patient-oriented perspective.
In: Dworkin RH, Breitbart W, eds. Psychosocial
Aspects of Pain: A Handbook for Health Care
38.

32 Pincus T, Sokka T. Quantitative target values of pre-
dictors of mortality in rheumatoid arthritis as possible
goals for therapeutic interventions: An alternative
approach to remission or ACR20 responses? J Rheu-

33 Sokka T, Pincus T. Markers for work disability in
22.

34 Clements PJ, Wong WK, Hurwitz EL, et al. The Disa
bility Index of the Health Assessment Questionnaire is
a predictor and correlate of outcome in the high-dose
versus low-dose penicillamine in systemic sclerosis

35 Fitzcharles MA, Costa DD, Poyhia R. A study of stan-
dard care in fibromyalgia syndrome: A favorable

36 Sadreddini S, Molaeefard M, Noshad H, Ardalan M,
Asadi A. Efficacy ofRaloxifien in treatment of fibromy-

37 Raftery G, Bridges M, Heslop P, Walker DJ. Are fibro-
myalgia patients as inactive as they say they are? Clin

38 Nantz E, Liu-Seifert H, Skljarevski V. Predictors of
premature discontinuation of treatment in multiple
43.

39 Clauw DJ, Mease P, Palmer RH, Gendreau RM, Wang
Y. Milnacipran for the treatment of fibromyalgia in
adults: A 15-week, multicenter, randomized, double-
blind, placebo-controlled, multiple-dose clinical trial.

40 Davis MJ, Addis ME. Predictors of attrition from
behavioral medicine treatments. Ann Behav Med

41 Dunbar J. Predictors of patient adherence: Patient
characteristics. In: Shumaker SA, Ockene JK, Riekert
KA, eds. The Handbook of Health Behavior Change.
New York: Springer Publishing Company; 2008:348–
60.

42 van Abben R, Van Wilgen CP, van der Schans CP,
Van Ittersum MW. Patients with more severe symp-
toms benefit the most from an intensive multimodal
programme in patients with fibromyalgia. Disabil


