Early Prediction of Fluoxetine Response for Han Chinese Inpatients With Major Depressive Disorder

Ching-Hua Lin, MD,† Hsien-Yuan Lane, MD, PhD,‡§ Cheng-Chung Chen, MD, PhD,* Suh-Hang Hank Juo, MD, PhD,||¶# and Cheng-Fang Yen, MD, PhD**††*

Abstract: The onset of antidepressant action is vital clinically. This study aimed to testify whether early symptom improvement can predict eventual treatment response at week 6 among depressive hospitalized patients taking fluoxetine. One hundred thirty-one hospitalized patients with major depressive disorder received 20 mg/d of fluoxetine for 6 weeks. Symptom severity was assessed by the 17-item Hamilton Depression Rating Scale (HAMD-17) at weeks 0, 1, 2, 3, 4, and 6. Stable response was defined as a reduction of 50% or more in the HAMD-17 total score at weeks 4 and 6 of treatment. Receiver operating characteristic curve was used to determine the cutoff point of the percentage of HAMD-17 score reduction between stable responders and nonresponders at weeks 1, 2, 3, and 4. At weeks 1, 2, 3, and 4, HAMD-17 score reductions of 25%, 39%, 43%, and 50% seemed to be the optimal cutoff points for predicting eventual response. They provided a sensitivity of 78%, 86%, 91%, and 93% and a specificity of 61%, 74%, 76%, and 92%. The percentage of HAMD-17 reduction at week 4 excellently predicted final response at week 6. Patients with less than a 50% symptom reduction during the first 4 weeks of treatment are unlikely to reach a final stable response. Whether this model can be applied to establish a prediction system for other antidepressants or for outpatients warrants further research.

Key Words: major depressive disorder, fluoxetine, early prediction model

(*J Clin Psychopharmacol* 2011;31: 187–193)

Major depressive disorder is common, and is associated with significant morbidity, mortality, and economic cost.[1,2] Antidepressants, such as selective serotonin reuptake inhibitors (SSRIs), are effective in the treatment of acute major depressive disorder.[3] Albeit continuing to develop in recent decades,[4,5] antidepressant medications still have a delayed onset of action.[6–9] Patients taking antidepressants often require a number of weeks to improve clinically.[10–14] For example, an earlier fixed-dose study[13] has suggested that early nonresponders to 6-week 20 mg/d of fluoxetine should continue to receive the same or higher dose of fluoxetine for additional weeks. The Sequenced Treatment Alternatives to Relieve Depression trial[15] reported that, of patients who ultimately showed clinical response when treated with open-label citalopram for 12 weeks, 56% first achieved response after 8 or more weeks, whereas 40% of patients who ultimately remitted first achieved remission after 8 or more weeks. On the basis of neurobiological studies, a delayed onset to antidepressants is often linked to the time taken for a variety of adaptive neurobiological changes to occur, for example, desensitization of serotonin 1A receptors and expression of neurotrophic factors such as brain-derived neurotrophic factor.[16–18]

Therefore, most guidelines for investigating the efficacy of antidepressants in treating the acute phase of major depressive disorder have recommended that such studies should have a duration of at least 4 weeks.[3,6,19–22]

The delayed onset of action is detrimental for patients with major depressive disorder. It increases patients' suffering and may contribute to hopelessness and treatment nonadherence.[23] On the other hand, several studies with virtually all groups of antidepressants also suggest that a treatment response can be observed within the first 2 weeks of treatment.[12,24–30] For example, Szegedi et al[30] found that improvement (defined as reduction in the 17-item Hamilton Depression Rating Scale [HAMD-17] score of ≥20% from the baseline) in the first 2 weeks in depressed patients treated with either mirtazapine or paroxetine was highly predictive of a positive response after 6 weeks of treatment. They also concluded that lack of early improvement was predictive of lack of improvement after 6 weeks. Although the sensitivities (percent of true positives) were high (97% for mirtazapine, 91% for paroxetine) even at week 2 of treatment, the specificities (percents of true negatives) were low all over the treatment: 53% for mirtazapine and 50% for paroxetine at week 2, 42% for mirtazapine and 36% for paroxetine at week 3, and 35% for mirtazapine and 30% for paroxetine at week 4. Another recent meta-analysis study by Szegedi et al[28] also used improvement (defined as reduction in HAMD-17 score of ≥20% compared with the baseline) within the first 2 weeks of treatment to predict the end point (4-8 weeks) stable response. The results were similar to their previous study,[30] indicating that early improvement with antidepressants can predict subsequent outcome with a high degree of sensitivity (eg, 88% for SSRI), but a low degree of corresponding specificity (eg, 50% for SSRI). One reason the 2 aforementioned studies[28,30] with lower specificities may be that the criteria of improvement (ie, at least 20%, 25%, or 30% symptom reduction) reflect only relatively minor symptom changes. Another reason may be the placebo effect. Quitkin et al[31] have suggested that early improvements after the start of antidepressant treatment often result from the placebo effect. Low specificities denoted that false-positive rates (1 − specificity) were high, and high false-positive rates cannot identify nonresponders earlier. Therefore, it is difficult to avoid unnecessarily continuing treatment with patients who would ultimately not respond. For early prediction of

From the *Kai-Suan Psychiatric Hospital, Kaohsiung; †Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung; ‡Institute of Clinical Medical Science and §Department of Psychiatry, China Medical University Hospital, Taichung; and ||Department of Medical Research, Kaohsiung Medical University Hospital, ¶Department of Medical Genetics, College of Medicine, #Center of Excellence for Environmental Medicine, **Department of Psychiatry, Faculty of Medicine, College of Medicine, and ††Department of Psychiatry, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan.

Received June 5, 2010; accepted after revision January 14, 2011.

Reprints: Cheng-Fang Yen, MD, PhD, Department of Psychiatry, Kaohsiung Medical University Hospital, 100 Tzyou 1st Rd, Kaohsiung City, Taiwan 807 (e-mail: chfaye@cc.kmu.edu.tw).

This study was funded by the Kai-Suan Psychiatric Hospital in 2007–2009, and National Science Council, Taiwan (NSC-97-2314-B-039-006-MY3 and NSC-98-2627-B-039-001), National Health Research Institutes, Taiwan (NHRI-EX-100-9904NI), Taiwan Department of Health Clinical Trial, and Research Center of Excellence (DOH100-TD-B-111-004).

ISSN: 0271-0749
DOI: 10.1097/JCP.0b013e318210856f

psychotropic response, high specificity is more important than high sensitivity in making treatment decisions as to whether the medication should be changed.[32]

When should we change the treatment after poor response? To increase the specificity of the prediction model, this study aimed to determine the optimal cutoff point of early symptom improvement for predicting eventual response/nonresponse at week 6 among inpatients with major depressive disorder taking fluoxetine, a widely used SSRI.

METHODS

Patients

The study was approved by Kai-Suan Psychiatric Hospital's institutional review board and conducted in accordance with Good Clinical Practice procedures and the current revision of the Declaration of Helsinki (project number: KSPH-2007-16). This study was also registered on Clinical.trials.gov (identifier number: NCT01075529).

Patients were recruited from Kai-Suan Psychiatric Hospital, a major psychiatric center in Taiwan, between May 2007 and February 2010. All newly hospitalized patients with major depressive disorder for immediate treatment were screened and evaluated by experienced psychiatrists. The *Structured Clinical Interview for Diagnostic* and *Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*)[33] criteria were used in ensuring the accuracy of the diagnosis. Han Chinese patients in Taiwan were enrolled in this study if they (1) were physically healthy and had all laboratory parameters within normal limits, (2) were aged 18 to 70 years, (3) satisfied *DSM-IV* criteria for major depressive disorder, (4) had a HAMD-17[34] score of 18 or higher and a Clinical Global Impression of Severity (CGI-S)[35] score of 4 or higher at baseline, (5) had no *DSM-IV* diagnosis of substance abuse or dependence (including alcohol) within the past 6 months, and (6) gave written informed consent to participate in the study after a full explanation of the study's aims and procedures. Patients excluded from this study were (1) those with a history of serious adverse reaction to fluoxetine or a history of epilepsy or organic mental disorders, (2) those with psychotic depression or bipolar I and II disorder, (3) those with schizophrenia or any other psychotic disorder, (4) those who were serious suicidal risks, (5) those with severe cognitive impairment, (6) women who were pregnant or at risk for pregnancy or lactation, (7) those patients initiating or stopping formal psychotherapy within 6 weeks before enrollment, or (8) those who have a history of poor response to fluoxetine or previously received electroconvulsive therapy.

Study Design and Procedures

After a washout period of at least 72 hours, patients received open-labeled fluoxetine treatment at a fixed dose of 20 mg daily for 6 weeks. Benzodiazepine (≤4 mg of lorazepam equivalent) or trazodone (≤100 mg) was allowed as needed at bedtime for insomnia. No other psychotropic agents were used.

Symptom severity was assessed at baseline and again at weeks 1, 2, 3, 4, and 6 by trained and experienced psychiatrists using HAMD-17. An intraclass correlation coefficient of 0.95 was obtained between the raters. To maintain high interrater reliability and prevent rater drift, raters met at least once a month for training and reliability retesting. The research psychiatrists who conducted the clinical ratings did not know the detailed study design or the responder versus nonresponder status of patients as defined during the study. Adverse effects were evaluated at each visit by the UKU Side Effect Rating Scale,[36] with scores ranging from 0 (none) to 3 (severe). A score of 1, 2, or 3 on any UKU item that first occurred or worsened during treatment indicated "cases" of adverse events. UKU was administered at baseline and at weeks 1, 2, 3, 4, and 6.

Improvement in depression severity was evaluated by the percentage of HAMD-17 score reduction from the baseline to each of the postbaseline assessment periods up to week 6 (ie, weeks 1, 2, 3, 4, and 6). This percentage of HAMD-17 score reduction from baseline to end point was calculated by the following formula:

$$[(\text{Baseline HAMD-17-Exit HAMD-17}) / \text{Baseline HAMD-17}] \times 100\%.$$

Stable response was defined as a reduction of 50% or more of the HAMD-17 score at weeks 4 and 6 of treatment.[30] Thus, for the present study, response status was a dichotomous outcome variable operationally defined as "stable response" or "nonresponse" of reduction in depression severity. Possible predictors were the percentage of HAMD-17 score reduction from baseline at weeks 1, 2, 3, and 4, respectively.

Statistical Analyses

Initially, stable responders and nonresponders at week 6 were compared in demographic data (sex, age), age at onset, number of previous episodes, baseline CGI-S score, baseline

TABLE 1. Clinical Characteristics Between Dropout Patients and Completers

	Dropout Patients (n = 19)		Completers (n = 112)		Analysis
Variables	**n**	**%**	**n**	**%**	***P****
Sex					0.14
Male	7	36.8	24	21.4	
Female	12	63.2	88	78.6	
	Mean	**SD**	**Mean**	**SD**	***P*†**
Age, y	42.9	11.0	45.6	11.0	0.32
Age at onset, y	34.0	10.6	39.5	11.8	0.06
No. previous episodes	2.2	1.6	2.4	2.0	0.67
Baseline CGI-S score	6.2	0.7	6.2	0.7	0.85
Baseline HAMD-17 score	29.5	5.2	31.6	6.7	0.19

*Pearson χ^2 test.

†Independent *t* test.

TABLE 2. Clinical Characteristics Comparing Stable Responders and Nonresponders After 6 Weeks of Treatment

	Stable Responders* (n = 58)		Nonresponders (n = 54)		Analysis
Variables	**n**	**%**	**n**	**%**	***P*†**
Sex					0.79
Male	13	22.4	11	20.4	
Female	45	76.6	43	79.6	
	Mean	**SD**	**Mean**	**SD**	***P*‡**
Age, y	46.3	10.3	44.9	11.8	0.48
Age at onset, y	40.5	12.4	38.4	11.1	0.34
No. previous episodes	2.3	2.1	2.7	1.7	0.28
Baseline CGI-S score	6.2	0.7	6.3	0.6	0.16
Baseline HAMD-17 score	31.0	7.3	32.4	5.9	0.27

*Patients with a reduction of 50% or more of the HAMD-17 score at weeks 4 and 6 of treatment.
†Pearson χ^2 test.
‡Independent *t* test.

HAMD-17 score, and percentages of HAMD-17 score reduction at weeks 1, 2, 3, and 4. Age at onset was regarded as age at the first major depressive episode.

Second, if the percentages of HAMD-17 score reduction at week 1, 2, 3, or 4 were reliable predictors after statistical analysis, they were entered in a receiver operating characteristic (ROC) curve. The ROC curve was used to determine the best cutoff point of predictor between the stable responders and nonresponders to maximize both the sensitivity and the specificity of the predictor variable so that false-positive and false-negative rates could be minimized. The area under the ROC curve (AUC) is a parameter used to quantify the ability or accuracy of the test in identifying stable responders from nonresponders. In practice, an AUC generally falls somewhere between 0.50 and 1.

Pearson χ^2 test or the Fisher exact test was used to compare categorical variables; the independent *t* test was used for continuous variables. All tests were 2-tailed, and significance was defined as an α of less than 0.05.

Effect-size (*d*) statistics were calculated to ascertain the degree to which the resulting changes in symptom severity were clinically recognizable. Effect size was calculated using the formula: d = (baseline mean − end point mean) / pooled SD. According to Cohen,[37] a d value of 0.20 indicates a small effect size, 0.50 a medium effect size, and 0.80 a large effect size. The large effect sizes demonstrate clinically relevant improvement at the end point.

All data were processed by SPSS version 17.0 for Windows (SPSS, Inc, Chicago, Ill) and MedCalc (MedCalc Software, Belgium). The MedCalc software is a program that implements several statistical procedures, including ROC analysis.

RESULTS

Patients

A total of 131 acutely ill inpatients with major depressive disorder were enrolled. Of the 131 patients, 112 (85.5%)

TABLE 3. Adverse Events Occurring in at Least 20% of Patients in Either Group

	Stable Responders (n = 58)		Nonresponders (n = 54)		Analysis
Adverse Events	**n**	**%**	**n**	**%**	***P***
At least 1 adverse event	52	89.7	53	98.1	0.07*
Concentration difficulties	9	15.5	17	31.5	0.046†‡
Asthenia/increased fatigability	10	17.2	13	24.1	0.37†
Failing memory	7	12.1	11	20.4	0.23†
Tension	4	6.9	13	24.1	0.01*‡
Reduced duration of sleep	2	3.4	13	24.1	<0.01*‡
Accommodation disturbances	8	13.8	15	27.8	0.07†
Reduced salivation	15	25.9	26	48.1	0.01†‡
Constipation	16	27.6	12	22.2	0.51†
Polyuria	15	25.9	18	33.3	0.39†
Orthostatic dizziness	20	34.5	26	48.1	0.14†
Palpitation	8	13.8	17	31.5	0.03‡†
Headache	6	10.3	12	22.2	0.09†

*Fisher exact test.
†Pearson χ^2 test.
‡Statistically significant.

TABLE 4. Early Predictors at Weeks 1, 2, 3, and 4 for Stable Response (≥50% Reduction in HAMD-17 at Weeks 4 and 6)

	Stable Responders (n = 58)		Nonresponders (n = 54)			
Early Predictor	**Mean**	**SD**	**Mean**	**SD**	**Effect Size (*d*)***	***P*†**
Percentage of HAMD-17 score reduction at week 1	41.2	19.0	22.2	18.3	1.02	<0.001‡
Percentage of HAMD-17 score reduction at week 2	57.4	19.0	30.2	17.6	1.49	<0.001‡
Percentage of HAMD-17 score reduction at week 3	64.2	17.6	31.6	18.6	1.81	<0.001‡
Percentage of HAMD-17 score reduction at week 4	71.1	13.4	32.0	16.4	2.61	<0.001‡

*d = 0.20 is defined as a small effect size; d = 0.50, a medium effect size; and d = 0.80, a large effect size.

†Independent t test.

‡Statistically significant.

completed the 6-week trial of fluoxetine (completers). The remaining 19 did not complete the trial owing to the lack of efficacy (n = 3), premature discharge (n = 14), or withdrawal of consent (n = 2). No patient withdrew from the study because of adverse events. The dropout patients (n = 19) and the completers (n = 112) were comparable for sex, age, age at onset, number of previous episodes, baseline CGI-S scores, and baseline HAMD-17 scores (Table 1).

Stable Response

Of the 112 completers, 51.8% (n = 58) of the subjects were classified as stable responders after 6 weeks. Stable responders and nonresponders did not differ in sex, age, age at onset, number of previous episodes, baseline CGI-S scores, and baseline HAMD-17 scores (Table 2).

Adverse Events

Adverse events occurring at an incidence of 20% or higher in any treatment group are shown in Table 3. Nonresponders were more likely to experience concentration difficulties, tension, reduced duration of sleep, reduced salivation, and palpitation. No severe adverse effects (score = 3) in any of the UKU items were noted in any of our patients.

Early Prediction

Table 4 demonstrates that the percentages of HAMD-17 score reduction were significantly different at each of the assessment weeks for stable responders and nonresponders. As early as week 1, stable responders experienced significantly greater percentages of HAMD-17 score reduction. The large effect sizes demonstrate clinically relevant improvement at weeks 1, 2, 3, and 4. This analysis indicated that the percentages of HAMD-17 score reduction at week 1, 2, 3, or 4 were reliable predictors of stable responders.

Finally, ROC analysis was used to determine the cutoff point of score changes as reliable predictors by plotting the proportion of true-positive results (sensitivity) versus the proportion of false-positive results (1 − specificity). At week 1, a HAMD-17 score reduction of 25% seemed to be the optimal cutoff point for predicting eventual response, providing a sensitivity of 78%, a specificity of 61%, and a predictive power (= number of true positives plus number of true negatives divided by total number of patients) of 70%. At week 2, a HAMD-17 score reduction of 39% seemed to be the optimal cutoff point for predicting eventual response, providing a sensitivity of 86%, a specificity of 74%, and a predictive power of 80%. At week 3, a HAMD-17 score reduction of 43% seemed to be the optimal cutoff point for predicting eventual response, providing a sensitivity of 91%, a specificity of 76%, and a predictive power of 84%. At week 4, a HAMD-17 score reduction of 50% seemed to be the optimal cutoff point for predicting eventual response, providing a sensitivity of 93%, a specificity of 92%, and a predictive power of 93% (Table 5). The percentage of HAMD-17 reduction at week 4 predicted nonresponse at week 6 better than the percentage of HAMD-17 reductions at week 1, 2, or 3 because the sensitivity and specificity values were generally higher. The ROC curves at weeks 1, 2, 3, and 4 are presented in Figure 1.

DISCUSSION

The main finding of this study is that a HAMD-17 score reduction of 25% or higher at week 1, 39% or higher at week 2, 43% or higher at week 3, and 50% or higher at week 4 correctly identified an ultimate stable response at the end of the study in

TABLE 5. Prediction of Stable Response (≥50% HAMD-17 Score Reduction at Weeks 4 and 6) Using Percentage of HAMD-17 Score Reduction at Weeks 1, 2, 3, and 4: ROC Analysis*

Early Predictors	**Percentage of Score Reduction Used as Cutoff Point, %**	**Sensitivity, %**	**Specificity, %**	**Predictive Power, %**	**AUC, %**
Percentage of HAMD-17 score reduction at week 1	25	78	61	70	77
Percentage of HAMD-17 score reduction at week 2	39	86	74	80	86
Percentage of HAMD-17 score reduction at week 3	43	91	76	84	91
Percentage of HAMD-17 score reduction at week 4	50	93	92	93	98

*ROC indicates receiver operating characteristic, for determining the cutoff point of score change as the predictor by plotting the proportion of true-positive results (sensitivity) versus the proportion of false-positive results (1 − specificity).

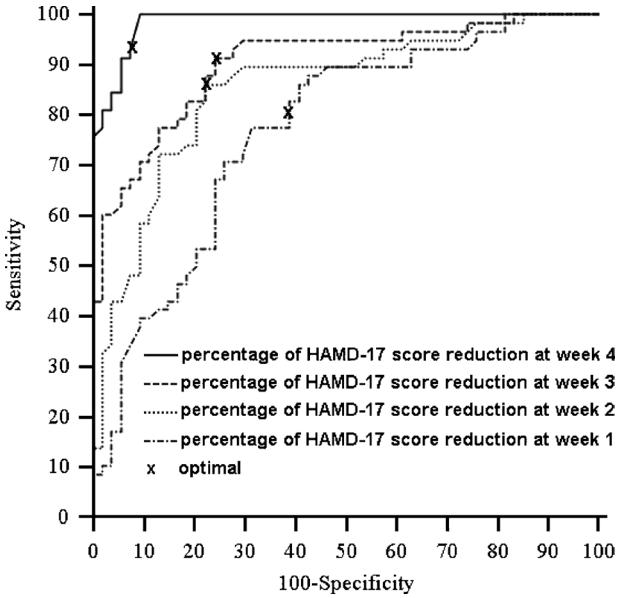

FIGURE 1. The ROC curves for stable responders versus nonresponders at weeks 1, 2, 3, and 4.

78%, 86%, 91%, and 93% of all patients. On the other hand, 61% of patients who displayed a less than 25% HAMD-17 score reduction at week 1, 74% of patients who displayed a less than 39% HAMD-17 score reduction at week 2, 76% of patients who displayed a less than 43% HAMD-17 score reduction at week 3, and 92% of patients who displayed a less than 50% HAMD-17 score reduction at week 4 were correctly identified as ultimate nonresponders. The results indicated that the percentage of HAMD-17 reduction at week 4 may be a better early predictor than that at week 1, 2, or 3. In clinical practice, however, a period of at least 4 weeks is worth attempting before any change in treatment should be considered for inpatients with a HAMD-17 score reduction less than 50%.

Like the studies of Szegedi et al,[28,30] if we use 20% improvement of HAMD-17 as the cutoff point, the sensitivity will be 93% and the specificity will be 37%. Of 112 patients, 87 will show 20% improvement of HAMD-17 score after 2 weeks of treatment. However, 33 of 87 patients will not respond after 6 weeks of treatment (ie, low specificity). This result is consistent with that of the studies by Szegedi et al[28,30] that very few patients who had not improved after 2 weeks became a stable responder after 4 weeks. However, a substantial number of patients who had shown early improvement did not become stable responders later, suggesting a limited specificity of the predictor. The current study suggests that ROC curve analysis can help solve this puzzling problem but after a longer treatment duration of 4 weeks.

Traditionally, response is frequently defined as a reduction of 50% or more of the HAMD-17 total score after 6 weeks of treatment. According to the traditional definition, a total of 66 (58.9%) of the 112 completers were classified as responders after 6 weeks of treatment. This response rate (58.9%) was similar to rates (50%–70%) found in placebo-controlled, 6 to 8 weeks, randomized, controlled trials with SSRIs.[23,38,39] There were no significant differences in sex, age, number of previous episodes, baseline CGI-S scores, and baseline HAMD-17 scores between responders and nonresponders (data not shown). The percentages of HAMD-17 score changes at weeks 1 ($P < 0.001$), 2 ($P < 0.001$), 3 ($P < 0.001$), and 4 ($P < 0.001$) were also reliable predictors of response. The cutoff points at which sensitivity and specificity were optimal were at percent changes of 25% (sensitivity, 71%; specificity, 59%), 39% (77%; 73%), 40% (83%; 74%), and 45% (92%; 87%), respectively.

If stable remission (defined as a HAMD-17 score ≤7 at weeks 4 and 6[28,30]) or remission (a HAMD-17 score ≤7 at week 6) were regarded as one of the outcome measures in this study, 26 (23.2%) of 112 patients became stable remitters, or 31 (27.7%) became remitters after 6 weeks of treatment. A HAMD-17 reduction of 63%, or 60% at week 4, seemed to be the optimal cutoff point for predicting ultimate stable remission or remission. It provided a sensitivity of 92% or 90% and a specificity of 83% or 81%, respectively. A percentage of HAMD-17 reduction at week 4, which yielded the optimal combined sensitivity and specificity, was better for predicting nonremitters at week 6 than those percentages at weeks 1, 2, or 3 (data not shown). Regardless of stable response, response, stable remission, or remission, decisions might be made at 4 weeks of treatment to determine whether a patient should be maintained on the initial drug or shifted to a new treatment.

Patients in the current study received the same fixed dose, 20 mg daily, of fluoxetine treatment. Earlier fixed-dose studies[40,41] have demonstrated that 20 mg of fluoxetine daily is the optimal dose for most patients and is associated with fewer and less severe adverse effects than higher doses. A meta-analysis study by Beasley et al[42] also found that fluoxetine therapy at 20 mg daily is a critical factor for adequate therapy and has good treatment tolerance. However, the rate and quality of response to fluoxetine are highly individualized.

Several strengths of this study could be addressed. First, the subjects were inpatients for immediate treatment. Hospitalized patients constituted only a small proportion of the patients in the studies.[43] As inpatients, they were carefully monitored, including symptom assessment, the development of adverse effects, and medical adherence. They also had the similar environmental conditions. Second, inpatients reflect greater severity of depression than outpatients. It has been reported that patients with more severe depression have less of a placebo effect.[44] Third, HAMD-17 was originally developed for inpatients.[34] It might be less, or even insufficiently sensitive, in detecting changes in depressive symptoms of minor severity.[12,15] Finally, unlike other studies that used last-observation-carried-forward analysis to account for missing data, we analyzed only the trial completers. The last-observation-carried-forward analysis, assuming that a subject's severity rating at the time of dropout would be the same as his or her rating at the end of the trial, could add a negative bias to the results across time.[29]

The limitations of this study included a relatively small, short-term, open-labeled treatment design, the use of a single antidepressant agent (fluoxetine), and having been conducted in only 1 psychiatric center. Because this was a relatively short-term trial, we did not know whether we would have had consistent findings for 8 or 12 weeks of treatment. The difficulty of a long-term study is that most of the inpatients enrolled do not agree to stay in hospital beyond 6 weeks. Although this was an open-labeled study, its goal was to early identify poor responders rather that to demonstrate treatment efficacy. Like other open trials,[45] the results could be generalized to a clinical setting, in which patients and clinicians both know the medication and both expect the outcomes. Further, the subjects comprised inpatients who had been hospitalized because of severe symptoms, severe functional impairment, or suicidal tendencies. Thus, inclusion of a placebo group had ethical concern. Moreover, it was also unclear if the conclusions could be generalized to other antidepressants or outpatients. However, a major difference in the

treatment between inpatients and outpatients is that medical adherence can be ascertained in hospitalization. If a patient has a good medical adherence after discharge, we assume that the conclusion of our study could be generalizable to him or her. On the other hand, the patients in the current study represent a relatively severely ill population who needs hospitalization. Therefore, whether the finding could be fully extrapolated to patients with less severity requires further studies.

Further studies, preferably involving other antidepressants, larger inpatient or outpatient groups from multicenters, and duration of longer than 6 weeks, are needed to better determine the predictive value of initial symptom change for ultimate treatment response.

AUTHOR DISCLOSURE INFORMATION

The authors declare no conflicts of interest.

REFERENCES

1. Greenberg PE, Stiglin LE, Finkelstein SN, et al. The economic burden of depression in 1990. *J Clin Psychiatry.* 1993;54:405–418.
2. Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. *J Affect Disord*. 2002;72:227–236.
3. Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. *Am J Psychiatry.* 2000;157(suppl 4):1–45.
4. Katz MM, Koslow SH, Maas JW, et al. The timing, specificity and clinical prediction of tricyclic drug effects in depression. *Psychol Med*. 1987;17:297–309.
5. Derivan AT. Antidepressants: can we determine how quickly they work? Issues from the literature. *Psychopharmacol Bull*. 1995;31:23–28.
6. Anderson IM, Ferrier IN, Baldwin RC, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. *J Psychopharmacol*. 2008;22:343–396.
7. Montgomery SA. Fast-onset antidepressants. *Int Clin Psychopharmacol*. 1997;12(suppl 3):1–5.
8. Quitkin FM, Rabkin JG, Ross D, et al. Duration of antidepressant drug treatment. What is an adequate trial? *Arch Gen Psychiatry.* 1984;41:238–245.
9. Harmer CJ, Goodwin GM, Cowen PJ. Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *Br J Psychiatry.* 2009;195:102–108.
10. Quitkin FM, McGrath PJ, Stewart JW, et al. Chronological milestones to guide drug change. When should clinicians switch antidepressants? *Arch Gen Psychiatry.* 1996;53:785–792.
11. Donovan SJ, Quitkin FM, Stewart JW, et al. Duration of antidepressant trials: clinical and research implications. *J Clin Psychopharmacol*. 1994;14:64–66.
12. Nierenberg AA, Farabaugh AH, Alpert JE, et al. Timing of onset of antidepressant response with fluoxetine treatment. *Am J Psychiatry.* 2000;157:1423–1428.
13. Wernicke JF, Dunlop SR, Dornseif BE, et al. Fixed-dose fluoxetine therapy for depression. *Psychopharmacol Bull*. 1987;23:164–168.
14. Frazer A, Benmansour S. Delayed pharmacological effects of antidepressants. *Mol Psychiatry.* 2002;7(suppl 1):23–28.
15. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006;163:1905–1917.
16. Blier P, de Montigny C, Chaput Y. Modifications of the serotonin system by antidepressant treatments: implications for the therapeutic response in major depression. *J Clin Psychopharmacol*. 1987;7(suppl 6):24–35.
17. Duman RS, Heninger GR, Nestler EJ. A molecular and cellular theory of depression. *Arch Gen Psychiatry.* 1997;54:597–606.
18. Hyman SE, Nestler EJ. Initiation and adaptation: a paradigm for understanding psychotropic drug action. *Am J Psychiatry.* 1996;153:151–162.
19. Guidelines on psychotropic drugs for the EC. Committee for Proprietary Medicinal Products. European Economic Community. *Eur Neuropsychopharmacol*. 1994;4:61–77.
20. Reesal RT, Lam RW. Clinical guidelines for the treatment of depressive disorders. II. Principles of management. *Can J Psychiatry.* 2001;46(suppl):1:21–28.
21. Bauer M, Bschor T, Pfennig A, et al. World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Unipolar Depressive Disorders in Primary Care. *World J Biol Psychiatry.* 2007;8:67–104.
22. Qaseem A, Snow V, Denberg TD, et al. Using second-generation antidepressants to treat depressive disorders: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2008;149:725–733.
23. Sackeim HA, Roose SP, Lavori PW. Determining the duration of antidepressant treatment: application of signal detection methodology and the need for duration adaptive designs (DAD). *Biol Psychiatry.* 2006;59:483–492.
24. Katz MM, Tekell JL, Bowden CL, et al. Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. *Neuropsychopharmacology.* 2004;29:566–579.
25. Papakostas GI, Perlis RH, Scalia MJ, et al. A meta-analysis of early sustained response rates between antidepressants and placebo for the treatment of major depressive disorder. *J Clin Psychopharmacol*. 2006;26:56–60.
26. Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? A meta-analysis. *J Clin Psychiatry.* 2005;66:148–158.
27. Stassen HH, Angst J, Hell D, et al. Is there a common resilience mechanism underlying antidepressant drug response? Evidence from 2848 patients. *J Clin Psychiatry.* 2007;68:1195–1205.
28. Szegedi A, Jansen WT, van Willigenburg AP, et al. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. *J Clin Psychiatry.* 2009;70:344–353.
29. Taylor MJ, Freemantle N, Geddes JR, et al. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Arch Gen Psychiatry.* 2006;63:1217–1223.
30. Szegedi A, Muller MJ, Anghelescu I, et al. Early improvement under mirtazapine and paroxetine predicts later stable response and remission with high sensitivity in patients with major depression. *J Clin Psychiatry.* 2003;64:413–420.
31. Quitkin FM, Petkova E, McGrath PJ, et al. When should a trial of fluoxetine for major depression be declared failed? *Am J Psychiatry.* 2003;160:734–740.
32. Leucht S, Busch R, Kissling W, et al. Early prediction of antipsychotic nonresponse among patients with schizophrenia. *J Clin Psychiatry.* 2007;68:352–360.
33. American Psychiatric Association. *Structured Clinical Interview for DSM-IV.* Washington, DC: American Psychiatric Press; 1994.
34. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56–62.
35. Guy W. *ECDEU Assessment Manual for Psychopharmacology.* Washington, DC: DHEW Publication; 1976:76338.
36. Lingjaerde O, Ahlfors UG, Bech P, et al. The UKU Side Effect Rating Scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand Suppl*. 1987;334:1–100.

37. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
38. Davidson JR, Meoni P, Haudiquet V, et al. Achieving remission with venlafaxine and fluoxetine in major depression: its relationship to anxiety symptoms. *Depress Anxiety.* 2002;16:4–13.
39. Entsuah AR, Huang H, Thase ME. Response and remission rates in different subpopulations with major depressive disorder administered venlafaxine, selective serotonin reuptake inhibitors, or placebo. *J Clin Psychiatry.* 2001;62:869–877.
40. Schweizer E, Rickels K, Amsterdam JD, et al. What constitutes an adequate antidepressant trial for fluoxetine? *J Clin Psychiatry.* 1990;51:8–11.
41. Stokes PE. Fluoxetine: a five-year review. *Clin Ther.* 1993;15:216–243.
42. Beasley CM Jr, Nilsson ME, Koke SC, et al. Efficacy, adverse events, and treatment discontinuations in fluoxetine clinical studies of major depression: a meta-analysis of the 20-mg/day dose. *J Clin Psychiatry.* 2000;61:722–728.
43. Gram L. Fluoxetine. *N Engl J Med*. 1994;331:1354–1361.
44. Khan A, Leventhal RM, Khan SR, et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002;22:40–45.
45. Nierenberg AA, McLean NE, Alpert JE, et al. Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. *Am J Psychiatry.* 1995;152:1500–1503.