

Latency to Rapid Eye Movement Sleep as a Predictor of Treatment Response to Fluoxetine and Placebo in Nonpsychotic Depressed Outpatients

John H. Heiligenstein, Douglas E. Faries, A. John Rush, John S. Andersen, Atul C. Pande, Howard P. Roffwarg, David Dunner, J. Christian Gillin, Steven P. James, Henry Lahmeyer, John Zajecka, Gary D. Tollefson, and Dawn M. Gardner

Received September 1, 1993; revised version received January 7, 1994; accepted February 7, 1994.

Abstract. Fluoxetine and placebo were compared in 89 outpatients with major depression with ($n = 45$) or without ($n = 44$) a reduced or shortened rapid eye movement latency (SREML) (≤ 65 minutes) to determine whether rapid eye movement latency (REML) predicted placebo and/or antidepressant response. Men and women were stratified based on polysomnographic recordings and then randomly assigned to receive double-blind fluoxetine (20 mg/day) or placebo for 8 weeks after a 2-week, single-blind, placebo lead-in period. Fluoxetine-treated patients demonstrated a significantly greater reduction in the Hamilton Rating Scale for Depression total score and a significantly greater response rate than placebo-treated patients in both the SREML and the combined strata. Treatment differences in the non-SREML stratum were not statistically significant. Results supported REML as a predictor of placebo nonresponse but did not predict a differential fluoxetine response in patients with SREML compared with patients without SREML.

Key Words. Affective disorder, polysomnography, antidepressant response prediction.

John H. Heiligenstein, M.D., is Senior Research Physician; Gary D. Tollefson, M.D., Ph.D., is Executive Director; and Dawn M. Gardner, B.S., is Global Clinical Coordinator, Psychopharmacology Division, Lilly Research Laboratories; Douglas E. Faries, Ph.D., is Senior Research Statistician, Lilly Research Laboratories; John S. Andersen, Ph.D., is Research Scientist, Department of Decision Sciences, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN. A. John Rush, M.D., is Betty Jo Hay Distinguished Chair in Mental Health and Director, Mental Health Research Center, University of Texas, Southwestern Medical School, Dallas, TX. Atul C. Pande, M.D., who was formerly Assistant Professor, Department of Psychiatry, University of Michigan, Ann Arbor, MI, is currently Associate Research Physician, Psychopharmacology Division, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN. Howard P. Roffwarg, M.D., is Attending Psychiatrist, Terrell State Hospital, Dallas TX. David Dunner, M.D., is Professor, Department of Psychiatry; Director, Outpatient Psychiatry; and Vice-Chairman for Clinical Services, University of Washington Medical Center, Seattle, WA. J. Christian Gillin, M.D., is Professor of Psychiatry and Director, Mental Health Research Center, Department of Psychiatry, University of California at San Diego, La Jolla, CA. Steven P. James, M.D., is Clinical Assistant Professor, University of Southern California, and Medical Director, Yorba Hills Hospital, Yorba Linda, CA. Henry Lahmeyer, M.D., is Assistant Professor of Psychiatry, Northwestern University Medical Center, Chicago, IL. John Zajecka, M.D., is Assistant Professor, Department of Psychiatry; Clinical Director, Treatment Research Center; and Medical Director, Ambulatory Services, Rush Institute for Mental Well-Being, Rush-Presbyterian-St. Luke's Medical Center, Chicago, IL. (Reprint requests to Dr. J.H. Heiligenstein, Psychopharmacology Division, Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Drop Code 2128, Indianapolis, IN 46285, USA.)

Reduced or shortened rapid eye movement latency (SREML) is among the most promising of biologic predictors of response to somatic treatment for major depression. It is thought that depressed patients with SREML are less likely to respond to placebo than are depressed patients who are not characterized by SREML. In addition, depressed patients with SREML may have a higher response rate to tricyclic antidepressants than those without SREML.

SREML as a response predictor has been relatively ignored (Joyce and Paykel, 1989). In one of the few studies, Coble et al. (1979) reported on 12 inpatients with SREML who had either a primary or biologic depression according to the Schedule for Affective Disorders and Schizophrenia (SADS; Spitzer and Endicott, 1978) nomenclature. After a 1-week, single-blind, placebo lead-in period, followed by 4 weeks of active psychosocial intervention and placebo treatment, none of the patients showed a remission or partial remission. All 12 placebo-treated patients eventually required pharmacologic treatment to be discharged from the hospital. The authors concluded that patients with SREML during the course of a biologic depressive disorder did not respond to placebo combined with active psychosocial treatment.

In a second study, Svendsen and Christensen (1981) reported on two groups of patients in which SREML predicted clinical response to somatic therapy. First, in 13 inpatients with endogenous depression and SREML, all responded satisfactorily to electroconvulsive therapy or antidepressant drugs. Second, in a distinct group of 10 outpatients with nonendogenous depression, the three patients without a reduced REML (≥ 80 minutes) had a poor result with treatment whereas the seven patients with SREML (≤ 50 minutes) had a favorable response after 5 weeks of somatic treatment. The authors concluded that SREML before treatment was associated with responsiveness to antidepressant therapy. In a third study, Rush et al. (1985) compared amitriptyline with alprazolam in a group of inpatients and outpatients with SREML (defined as ≤ 65 minutes) who were evaluated using the Research Diagnostic Criteria (RDC; Spitzer et al., 1975) for nonpsychotic major depression. Of the 49 patients (17 inpatients and 32 outpatients) who completed 3 or more weeks of drug treatment, 79% of amitriptyline-treated patients met the authors' remission criteria, in contrast to 36% of alprazolam-treated patients. The results of this study suggested that patients with SREML were much more likely to respond to a conventional antidepressant than to a benzodiazepine.

In a subsequent study, Rush et al. (1989) investigated 46 outpatients with nonpsychotic major depression according to SADS-L (Endicott and Spitzer, 1978) criteria. The patients were stratified into two groups: those with and those without SREML. A REML of ≤ 65 minutes defined those patients with SREML. The patients were then randomly assigned to double-blind treatment with desipramine or amitriptyline. The authors found that 80% of patients with SREML responded to either of the active treatments, whereas only 50% of patients identified as having non-SREML met the authors' criteria for treatment response. The combined response rate was 61% for all amitriptyline-treated patients and 67% for all desipramine-treated patients. The authors concluded that sleep polysomnography may predict treatment responsiveness to tricyclic antidepressants in patients with nonpsychotic major depression.

The present multicenter study was designed to compare the efficacy of fluoxetine, a selective serotonin uptake inhibitor, and placebo in patients with major depression who were and were not characterized by SREML. It was hypothesized that placebo response would be lower and active drug response higher in patients with SREML than in patients with non-SREML.

Methods

Patients. Patients included males and females, 18-65 years of age, who were experiencing a major depressive episode diagnosed according to *DSM-III-R* criteria (American Psychiatric Association, 1987) with the exception that patients must have satisfied these criteria for a minimum of 1 month. Patients suffered either major depression with a single or recurrent episode or were diagnosed as bipolar, type II depressed phase, according to RDC (Spitzer et al., 1975). Patients with more than one depressive episode had a minimum euthymic interval of 10 weeks between the two most recent depressive episodes. Patients were also required to have a minimum score of 15 on the first 17 items of the 28-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960).

Patients were excluded if they were pregnant or lactating; displayed serious medical illness; were psychotic; had not responded in the past to three or more antidepressants at a therapeutic dose (200 mg imipramine equivalents) for at least 3 weeks; had a seizure after age 12; had an organic mental disorder or a substance use disorder, including alcohol, active within the past year; or had an antisocial personality disorder and/or a history of three or more suicide attempts/gestures without clear concurrent major depressive disorder, melancholic type. Also excluded were patients who had a history of multiple adverse drug reactions or allergy to fluoxetine; were being treated with any hypertensive other than a diuretic or calcium channel blocker; were taking any other psychotropic medication, with the exception of chloral hydrate; had taken fluoxetine within 12 weeks of the polysomnographic studies; had the potential need to use a monoamine oxidase inhibitor within 5 weeks of discontinuation of treatment; or were at a serious risk of suicide. In addition, patients were excluded who had a diagnosis of narcolepsy, sleep apnea, or periodic limb movements of sleep by history or by sleep-staging polysomnography (PSG); had an increased thyroid stimulating hormone level or were taking thyroid supplements; had a *DSM-III-R* diagnosis of generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, phobias, or posttraumatic stress disorder; had conditions or took medications that could possibly influence the PSG determination of a SREML; or were involved in ongoing psychotherapy.

Study Design. Patients were stratified as having SREML or non-SREML based on PSG recordings and were then randomly assigned to receive double-blind fluoxetine or placebo therapy. The design included an adaptive feature intended to minimize patient exposure to ineffective therapy by using early therapeutic outcome to adjust the random allocation scheme.

The study was divided into two study periods: a 2-week (14- to 18-day) single-blind, drug washout placebo lead-in period to eliminate placebo responders and to prepare patients for PSG studies followed by an 8-week double-blind study period. Patients who did not respond to placebo had sleep staging PSG on 2 consecutive nights. On the basis of the mean value of the two PSG recordings, patients were stratified into two groups, provided they continued to meet criteria as placebo nonresponders at visit 4 and their visit 3 urine drug screen was negative. One stratum comprised patients with major depression and SREML. The second stratum comprised patients with major depression and non-SREML. Following the single-blind placebo lead-in period, patients who qualified for randomization were assigned within each stratum to placebo or fluoxetine, 20 mg/day, for 8 weeks. With the exception of chloral hydrate, no additional psychoactive compounds were permitted. Fluoxetine and placebo were given in the morning. Patients were seen weekly. Medication compliance was assessed by a

capsule count (number of capsules dispensed minus the number returned) at each visit.

Study Procedures. PSG recordings were scored visually according to the criteria of Rechtschaffen and Kales (1968). REML was defined as the time in minutes between sleep onset and the first REM period minus any time of wakefulness or movement during that interval. Sleep onset was defined as the time point that marked the beginning of at least 10 consecutive minutes of EEG sleep, including Stage 1 sleep, with no more than 2 minutes of wakefulness or movement within the period. REM sleep onset was defined as the time at which the PSG recording indicated the beginning of more than 15 seconds of REM sleep during a 30-second recording epoch.

Patients were defined as having SREML if the mean of their REMLs from 2 nights of PSG recording was ≤ 65 minutes. If for a given patient, one REML value was < 65 minutes and the other was > 65 minutes, with a spread of > 20 minutes between the two, the REML for that patient was the shorter REML recorded.

Randomization of treatments was performed using an adaptive allocation scheme similar to that described by Wei and Durham (1978). The adaptive allocation scheme allowed for an increased probability that a patient would receive the treatment that was performing better within each stratum. Patients were randomly assigned a treatment using a computer-simulated urn containing balls. At the beginning of study, the urn for each stratum contained two balls, one labeled fluoxetine and the other, placebo. Each time a patient was to be randomly assigned, a ball was drawn from the appropriate urn (with replacement), the treatment label identified, and the patient assigned to receive the corresponding treatment. As patients in the study were determined to be successes or failures, the urn was updated. For instance, if a fluoxetine-treated patient was a success, a fluoxetine ball was added to the urn. If a fluoxetine-treated patient was a failure, then a placebo ball was added to the urn. The process was similar for placebo-treated patients. The expected result was that the urn would contain more balls for the successful treatment so that patients had a greater probability of receiving the more effective treatment. To guarantee a minimum number of patients on each treatment, the first six patients in each stratum were randomly assigned to treatment in a balanced fashion, not using the computerized urn.

To accelerate updating of the urn for each stratum, patients who had completed at least 3 weeks of double-blind therapy were considered a success if they had had 2 consecutive weeks with at least a 50% reduction from baseline in the HRSD-17 total score. This definition of success differed from the definition of responders used in the efficacy analysis. For the efficacy analysis, patients who had completed at least 3 weeks of double-blind therapy were considered responders to therapy if their end-point HRSD-17 total score decreased by at least 50% from baseline.

Efficacy was assessed by the HRSD-17 (Hamilton, 1960), the Montgomery-Åsberg Depression Rating Scale (MADRS; Montgomery and Åsberg, 1979), the Clinical Global Impression (CGI)-Severity and Improvement (Guy, 1976), and the Patient's Global Impression (PGI)-Improvement (Guy, 1976). The HRSD-17, MADRS, and CGI-Severity measures were done at each visit throughout the study, and the PGI was collected beginning with visit 2. In addition, the Agitation Rating Scale was scored by the investigator at each visit. This latter scale rates items relative to agitation based upon RDC (Spitzer et al., 1975). The Medical Outcomes Study-36 (Ware and Sherbourne, 1992), a patient-rated measure of functional well-being, was obtained at visits 4 and 12, or at the visit at which the patient was discontinued from study. The investigator-rated HRSD-17 served as the primary efficacy instrument.

Adverse events were elicited by questioning patients at each followup visit. Open questions by the interviewer, rather than a standardized form, were used to elicit adverse events. Only adverse events that first occurred or worsened after the start of double-blind therapy were analyzed. To ensure uniformity of language in this trial, events were reported using terminology derived from the Food and Drug Administration's COSTART thesaurus (US FDA, 1985).

Data Analyses and Statistical Methods. Treatment comparisons of response rates, the primary analysis, were performed for patients who had sufficient exposure to therapy (at least 3 weeks). In addition, treatment comparisons of remission rates (HRSD-17 total end-point score ≤ 6) were performed for patients who had at least 3 weeks of double-blind therapy. Change in HRSD-17 total scores from baseline to end point was also compared between treatments using a last-observation-carried-forward approach. All randomized patients with at least one postbaseline measurement were included in this analysis.

While inferences for most clinical trials are based upon p values from methods such as the χ^2 test and analysis of variance, these methods may not be appropriate for trials using an adaptive randomization scheme (Ware, 1989; Begg, 1990). Therefore, a Bayesian approach to the statistical analysis of the data was incorporated (Press, 1989).

Inferences were based upon the posterior probabilities that p_1 , the true probability of success on placebo, was greater than p_2 , the true probability of success on fluoxetine, given the data from the study. A posterior probability > 0.950 or < 0.050 was defined as a statistically significant result. For instance, a posterior probability < 0.050 was considered strong evidence that the response rate on fluoxetine was superior to that on placebo. A similar rule was used to compare treatments with respect to mean HRSD change. Noninformative prior distributions were used in all calculations. In this article, posterior probabilities are denoted by the term "Bayes p ."

Additional statistical methods were used to assess the robustness of the results. These methods included the randomization test, a weekly completers-only analysis, and several repeated measures methods (Wu and Bailey, 1988; Crowder and Hand, 1990; Dawson and Lagakos, 1991). Analyses were performed in Version 6 of SAS (SAS User's Guide, 1990) and Mathematica™ (Wolfram, 1988).

Results

One hundred sixty-four outpatients were initially enrolled and began the single-blind placebo lead-in period. Of these, 75 (45.7%) patients discontinued before randomization, 9 because of placebo response, 9 because of sleep disturbances (apnea, myoclonus), and 57 for other reasons (33, entry criteria not met; 3, positive urine drug screen; 13, patient decision; 2, noncompliance; 2, lost to followup; 2, adverse events; 1, lack of efficacy; and 1, entered after the cutoff date). A total of 89 patients were randomly assigned to double-blind therapy, 45 (50.6%) in the SREML stratum (fluoxetine, 23; placebo, 22) and 44 (49.4%) in the non-SREML stratum (fluoxetine, 23; placebo, 21). Table 1 shows demographic and baseline clinical characteristics of the randomly assigned patients by stratum and for the strata combined. Although a greater percentage of females were randomized to placebo in each stratum, no statistically significant differences in demographic or baseline characteristics were found. HRSD treatment differences were consistent across gender and strata.

Table 2 presents the probabilities of randomization to each treatment at four different time points. In each stratum, the probability that a patient would be randomized to fluoxetine was 0.50 at the beginning of the study. At the time the last patient in each stratum was enrolled, this probability had increased to 0.68 in the SREML stratum and 0.54 in the non-SREML stratum. Despite these probabilities, the treatment allocation remained nearly balanced in both strata. This was in part due to chance and in part to differences between the definition of success for updating the urn and the definition of a responder in the final efficacy analysis.

Table 1. Demographic and baseline characteristics

| Stratum/treatment | n | Age (yr) | | % | HRSD-17 total score | | REML ¹ | |
|------------------------|----|----------|------|------|---------------------|-----|-------------------|------|
| | | Mean | SD | | Mean | SD | Mean | SD |
| SREML | | | | | | | | |
| Fluoxetine | 23 | 44.0 | 10.7 | 60.9 | 21.5 | 3.7 | 63.5 | 21.2 |
| Placebo | 22 | 41.5 | 11.3 | 81.8 | 22.2 | 3.4 | 61.7 | 18.8 |
| Non-SREML | | | | | | | | |
| Fluoxetine | 23 | 39.1 | 10.9 | 65.2 | 20.6 | 3.2 | 90.6 | 27.9 |
| Placebo | 21 | 36.8 | 7.4 | 76.2 | 21.0 | 3.1 | 87.0 | 21.3 |
| Combined strata | | | | | | | | |
| Fluoxetine | 46 | 41.6 | 10.9 | 63.0 | 21.1 | 3.4 | 77.1 | 28.1 |
| Placebo | 43 | 39.2 | 9.8 | 79.1 | 21.6 | 3.3 | 74.1 | 23.6 |
| All patients | 89 | 40.4 | 10.4 | 70.8 | 21.3 | 3.3 | 75.6 | 25.9 |

Note. HRSD-17 = first 17 items of the Hamilton Rating Scale for Depression. Non-SREML = nonshortened rapid eye movement latency. REML = rapid eye movement latency. SREML = shortened rapid eye movement latency.

1. Mean REML from 2 nights.

Table 2. Adaptive randomization probabilities at four time points during the trial

| Stratum/treatment | Probability at time | | | |
|-------------------|------------------------|--------------------------------|---------------------------------|-----------------------|
| | First patient enrolled | One third of patients enrolled | Two thirds of patients enrolled | Last patient enrolled |
| SREML | | | | |
| Fluoxetine | 0.50 | 0.64 | 0.60 | 0.68 |
| Placebo | 0.50 | 0.36 | 0.40 | 0.32 |
| Non-SREML | | | | |
| Fluoxetine | 0.50 | 0.50 | 0.63 | 0.54 |
| Placebo | 0.50 | 0.50 | 0.37 | 0.46 |

Note. Non-SREML = nonshortened rapid eye movement latency. SREML = shortened rapid eye movement latency.

Table 3 shows study completion rates and reasons for early discontinuation. In both strata combined, 32 of 46 (69.6%) fluoxetine-treated and 33 of 43 (76.7%) placebo-treated patients completed the study. This difference was not statistically significant. Of the patients who discontinued early, only 7 of 46 (15.2%) fluoxetine-treated and 7 of 43 (16.3%) placebo-treated patients discontinued the study because of an adverse event (fluoxetine, 2; placebo, 0) or lack of efficacy (fluoxetine, 5; placebo, 7). Thirteen of 23 (56.5%) fluoxetine-treated patients and 16 of 22 (72.7%) placebo-treated patients in the SREML stratum completed the study. However, of the 10 fluoxetine-treated patients who discontinued the study before the last visit, 6 did so for reasons other than lack of efficacy or an adverse event (2, unavailable for followup; 2, patient decision; 2, protocol requirement). None of the fluoxetine-treated or placebo-treated patients in either group discontinued the study because of the occurrence of insomnia. In those with complaints of insomnia, three patients

Table 3. Study completion rates

| Variable | SREML | | | | Non-SREML | | | |
|--------------------|------------|--------|---------|--------|------------|--------|---------|---------|
| | Fluoxetine | | Placebo | | Fluoxetine | | Placebo | |
| | No. | (%) | No. | (%) | No. | (%) | No. | (%) |
| Randomly assigned | 23 | | 22 | | 23 | | 21 | |
| Completed 3 weeks | 20 | (87.0) | 21 | (95.5) | 21 | (91.3) | 21 | (100.0) |
| Completed 8 weeks | 13 | (56.5) | 16 | (72.7) | 19 | (82.6) | 17 | (81.0) |
| Discontinued for | | | | | | | | |
| Adverse event | 1 | (4.3) | 0 | (0.0) | 1 | (4.3) | 0 | (0.0) |
| Lack of efficacy | 3 | (13.0) | 5 | (22.7) | 2 | (8.7) | 2 | (9.5) |
| Other ¹ | 6 | (26.1) | 1 | (4.5) | 1 | (4.3) | 2 | (9.5) |

Note. Non-SREML = nonshortened rapid eye movement latency. SREML = shortened rapid eye movement latency.

1. Unavailable for followup (SREML: fluoxetine, 2; non-SREML: placebo, 1), patient decision (SREML: fluoxetine, 2; non-SREML: fluoxetine, 1; placebo, 1), and protocol requirement (SREML: fluoxetine, 2; placebo, 1).

(fluoxetine, 1; placebo, 2) were treated with chloral hydrate, the only prescription hypnotic permitted during the study. Each of the three patients completed the study, although only the fluoxetine-treated patient met response criteria. Results of the secondary analyses done to assess the effect of discontinuations on result robustness were all in agreement with the results from the primary analysis.

Table 4 presents the response and remission rates with fluoxetine and placebo in each stratum and in the combined strata. Twenty of 23 (87.0%) fluoxetine-treated

Table 4. Response and remission rates

| Stratum/treatment | No. treated ≥ 3 weeks | Measure | | | |
|-------------------|--------------------------|-----------------------|--------|------------------------|--------|
| | | Response ¹ | | Remission ² | |
| | | No. | (%) | No. | (%) |
| SREML | | | | | |
| Fluoxetine | 20 | 12 | (60.0) | 5 | (25.0) |
| Placebo | 21 | 7 | (33.3) | 3 | (14.3) |
| Bayes p^3 | | 0.047 | | 0.204 | |
| Non-SREML | | | | | |
| Fluoxetine | 21 | 13 | (61.9) | 11 | (52.4) |
| Placebo | 21 | 10 | (47.6) | 6 | (28.6) |
| Bayes p^3 | | 0.182 | | 0.062 | |
| Combined strata | | | | | |
| Fluoxetine | 41 | 25 | (61.0) | 16 | (39.0) |
| Placebo | 42 | 17 | (40.5) | 9 | (21.4) |
| Bayes p^3 | | 0.032 | | 0.042 | |

Note. Non-SREML = nonshortened rapid eye movement latency. SREML = shortened rapid eye movement latency. HRSD-17 = first 17 items of the Hamilton Rating Scale for Depression.

1. ≥ 50% decrease in end-point HRSD-17 total score after at least 3 weeks of double-blind therapy.

2. HRSD-17 total score ≤ 6 at end point after at least 3 weeks of double-blind therapy.

3. The value shown reflects the probability that the true fluoxetine rate is lower than the true placebo rate, given the results. A value > 0.950 or ≤ 0.050 is considered statistically significant.

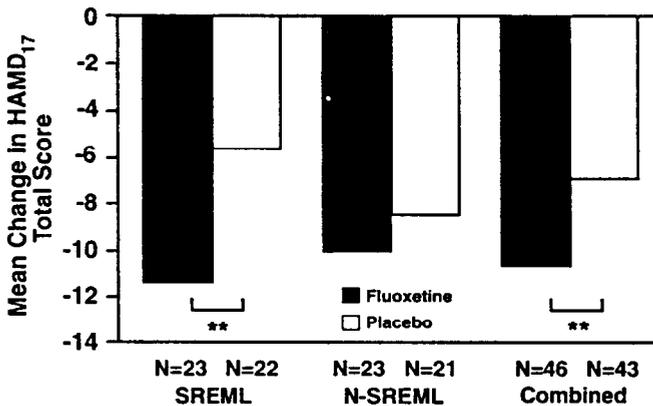
and 21 of 22 (95.5%) placebo-treated patients in the SREML stratum, and 21 of 23 (91.3%) fluoxetine-treated and 21 of 21 (100.0%) placebo-treated patients in the non-SREML stratum completed at least 3 weeks of treatment and were eligible for response and remission rate analyses. Fluoxetine produced a statistically significantly higher response rate (60%) than placebo (30%) in the SREML stratum (Bayes $p = 0.047$) and in the combined strata (Bayes $p = 0.032$). Although fluoxetine was numerically superior to placebo in achieving a remission in the SREML stratum, this difference was not statistically significant. Fluoxetine was significantly superior for remission rates in the total patient sample (39.0% vs. 21.4%, Bayes $p = 0.042$) and showed a trend to significance in remission rates in the non-SREML stratum (52.4% vs. 28.6%, Bayes $p = 0.062$). The results for an intent-to-treat analysis were similar.

Fig. 1 shows the baseline-to-end-point mean change in the HRSD-17 scores for patients by stratum and for the strata combined. As can be seen, fluoxetine was significantly superior to placebo in the SREML stratum (Bayes $p = 0.003$) and in the strata combined (Bayes $p = 0.008$), but not in the non-SREML stratum (Bayes $p = 0.269$). Fig. 2 shows the distribution of HRSD-17 total scores at end point (last visit carried forward) for patients in the SREML and non-SREML strata.

Two of the six investigational sites randomized three or fewer patients per REML stratum. The results were consistent across the other four investigational sites.

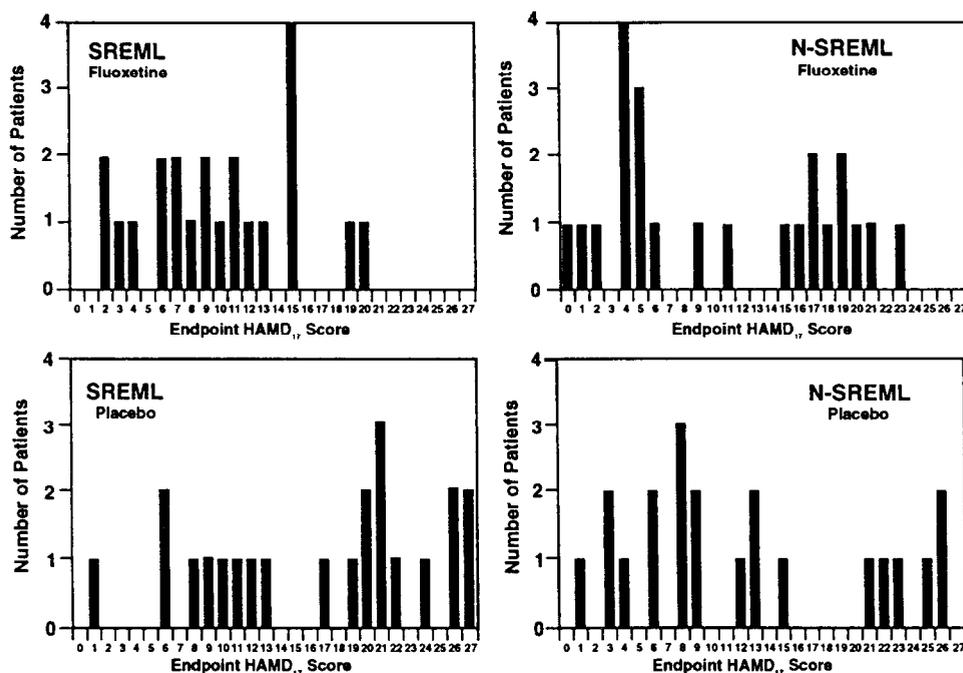
Fig. 3 illustrates the mean HRSD-17 change by week for the SREML stratum (top), non-SREML stratum (center), and combined strata (bottom). Fluoxetine demonstrated statistically significant differences from placebo in both the SREML stratum and the combined strata beginning with week 1 and continuing throughout the study with the exception that fluoxetine did not separate from placebo in the SREML stratum at week 4. There were no statistically significant differences at any week in the non-SREML stratum.

Fig. 1. Mean baseline-to-end-point change in HRSD-17 total scores by stratum and for the strata combined



SREML = shortened rapid eye movement latency stratum. N-SREML = nonshortened rapid eye movement latency stratum. HRSD = Hamilton Rating Scale for Depression. Asterisks denote a statistically significant treatment difference (** = Bayes $p < 0.01$).

Fig. 2. Distribution of HRSD-17 total scores at end point (last visit carried forward) for patients in the SREML and N-SREML strata



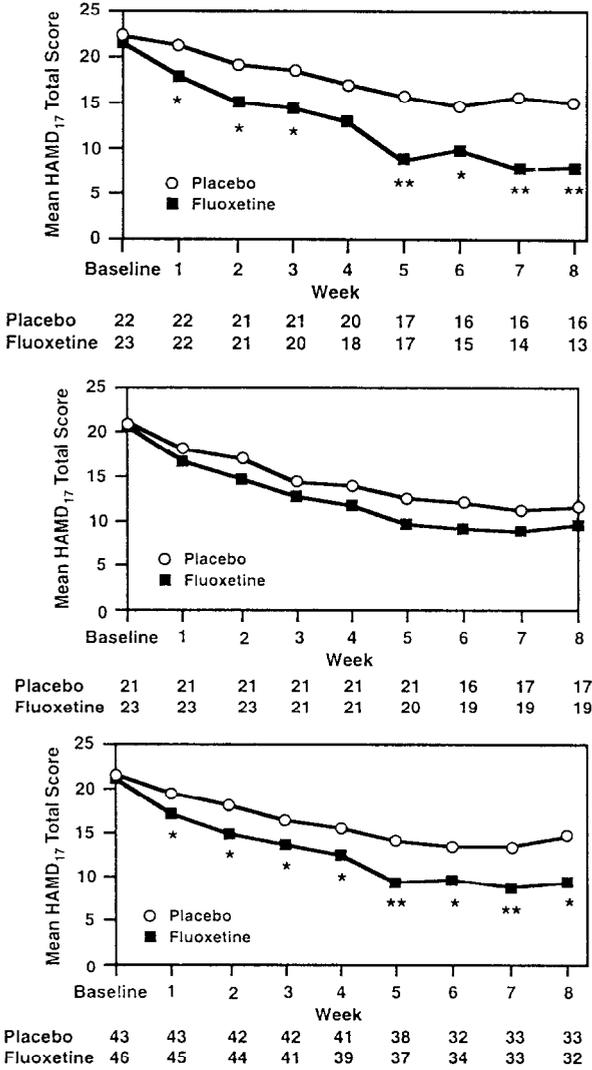
SREML = shortened rapid eye movement latency stratum. N-SREML = nonshortened rapid eye movement latency stratum. HRSD = Hamilton Rating Scale for Depression.

Table 5 summarizes changes in other measures of efficacy (MADRS total score and CGI-Severity). Significantly greater improvement in the MADRS total score (Bayes $p = 0.036$) and in CGI-Severity (Bayes $p = 0.016$) was observed with fluoxetine compared with placebo in the SREML stratum.

An analysis of adverse events showed that in the SREML stratum, asthenia occurred significantly more often with fluoxetine (17.4%) than with placebo (0.0%), while dyspepsia, flu syndrome, sinusitis, and urinary tract infection occurred significantly more often with placebo than fluoxetine. In the non-SREML stratum, back pain, constipation, and allergic reaction occurred significantly more often with placebo. In the combined strata, somnolence occurred significantly more often with fluoxetine (13.0%) than with placebo (2.3%), while constipation, sinusitis, and urinary tract infection occurred significantly more often with placebo than fluoxetine. No evidence of an increased incidence of treatment-emergent anxiety or nervousness attributable to fluoxetine was seen in this study. In the combined strata, 9 of 46 (19.6%) fluoxetine-treated patients experienced anxiety compared with 7 of 43 (16.3%) placebo-treated patients, whereas 4 (8.7%) fluoxetine-treated patients reported nervousness compared with 7 (16.3%) placebo-treated patients.

No suicide attempts were reported in this trial. To compare further the treatments with respect to effect on suicidality, the suicidality items from the HRSD (item 3)

Fig. 3. Mean weekly change in HRSD-17 total score for patients in the shortened rapid eye movement latency (SREML) stratum (top), non-shortened rapid eye movement latency (non-SREML) stratum (center), and combined strata (bottom).



SREML = shortened rapid eye movement latency stratum. N-SREML = nonshortened rapid eye movement latency stratum. HRSD = Hamilton Rating Scale for Depression. Asterisks denote a statistically significant treatment difference (* = Bayes $p < 0.05$; ** = Bayes $p < 0.01$). The numbers shown for placebo and fluoxetine indicate the sample size at each week.

Table 5. Summary of change in secondary efficacy measures¹

| Measure/ Variable | SREML | | Non-SREML | | Combined strata | |
|----------------------|------------------------|---------------------|------------------------|---------------------|------------------------|---------------------|
| | Fluoxetine (n = 22) | Placebo (n = 22) | Fluoxetine (n = 23) | Placebo (n = 21) | Fluoxetine (n = 45) | Placebo (n = 43) |
| MADRS total score | | | | | | |
| Baseline | 25.4±5.9 | 27.0± 5.9 | 26.2± 5.2 | 25.7± 5.6 | 25.8± 5.5 | 26.3± 5.8 |
| Change | -12.7±9.8 | -7.1±10.1* | -12.4±11.3 | -11.0±11.2 | -12.6±10.5 | -9.0±10.7 |
| CGI-Severity | | | | | | |
| Baseline | 4.0±0.5 | 4.2± 0.5 | 3.9± 0.5 | 4.0± 0.6 | 3.9± 0.5 | 4.1± 0.5 |
| Change | -1.5±1.0 | -0.7± 1.3* | -1.3± 1.4 | -1.4± 1.5 | -1.4± 1.2 | -1.1± 1.4 |

Note. SREML = shortened rapid eye movement latency. Non-SREML = nonshortened rapid eye movement latency. MADRS = Montgomery-Åsberg Depression Rating Scale. CGI = Clinical Global Impression.

*Statistically significant treatment difference, Bayes $p < 0.050$ or > 0.950 .

1. Values shown are mean ± standard deviation.

and the MADRS (item 10) were analyzed. Fluoxetine-treated patients demonstrated a significantly greater reduction in HRSD item 3 and MADRS item 10 scores than did placebo-treated patients in both the SREML stratum (Bayes $p = 0.004$) and in the combined strata (Bayes $p = 0.003$). Two cases of emergent suicidal ideation (an increase in baseline HRSD item 3 score from 0 or 1 to a score of 3 or 4 at any time during double-blind therapy) were observed, both in placebo-treated patients in the SREML stratum.

Comment

In this study, beginning at week 1 and continuing throughout the study, a robust and statistically significant separation of fluoxetine compared with placebo was demonstrated on the HRSD-17 total score in both the patients with SREML and the combined sample, demonstrating that the therapeutic benefits of fluoxetine were evident as early as the first week of therapy. A major finding of this study, however, was that SREML predicted relatively poor response to placebo compared with fluoxetine in patients with major depression. This finding would be expected based on results of previous studies (Coble et al., 1979; Svendsen and Christensen, 1981; Rush et al., 1985). Although there was no statistically significant difference in response rates between fluoxetine-treated and placebo-treated patients in the non-SREML group, the design of this study was not optimal to show a statistically significant treatment difference in patients with non-SREML.

Although statistically inferior to fluoxetine, placebo treatment performed better in the patients with SREML in our study than might have been expected from the work of Coble et al. (1979), whereas active treatment performed less well in the patients with SREML than might have been predicted from the work of Rush et al. (1985). However, direct comparisons may be limited by differences in study design, investigators, and methodology. Ours was a longer (8-week) multicenter study in outpatients with major depression as diagnosed by *DSM-III-R* criteria. Coble et al. (1979) and Rush et al. (1985, 1989) conducted shorter (4-week and 6-week, respectively) single-center studies with patients diagnosed by structured interview

using the SADS or SADS-Lifetime version. In addition, Coble et al. (1979) studied inpatients, while Rush et al. (1985) included both inpatients and outpatients in their first study. Since inpatients with major depression may be expected to be more severely ill, it would not be surprising to find a lower placebo response rate and a higher active drug response rate among them. Conversely, our outpatient, less severely depressed sample could have a trend toward narrower differentiation between active and placebo treatments. Moreover, our protocol did not allow the fluoxetine dose to be increased beyond 20 mg per day, thus eliminating the possibility of improvement with a higher dose in partial responders and nonresponders.

It is interesting to note that even though our sample was drawn from an outpatient pool, slightly more than 50% of the patients had SREML. This underscores the fact that careful clinical screening of outpatients can provide a homogeneous subgroup with a biological marker of major depression. We are conducting further analyses of the clinical correlates of this sample and will report them separately.

The adaptive randomization scheme was successful at increasing the probability, within the SREML stratum, that patients would receive fluoxetine, the treatment that performed better in this stratum. However, because of both chance and the definition of success used to accelerate updating of the computerized arm, treatment allocation remained balanced even in the SREML stratum. The logistics of performing a multicenter, double-blind adaptive trial proved to be feasible. The ethical advantage offered by adaptive trials makes them worth consideration in future studies.

While our findings are encouraging with respect to a possible biological marker for depression, their significance for a practicing clinician is less clear. This is especially so since fluoxetine was found to be an approximately equally effective treatment in both the SREML and the non-SREML strata. The use of REML as a "test" to predict who may benefit from treatment with a selective serotonin uptake inhibitor—in this case, fluoxetine—is not supported by our findings, but our data do suggest REML is a tool to predict who may or may not respond to placebo. Although of little practical utility for a clinician, this is of great interest to treatment researchers.

Acknowledgments. The research reported was supported in part by a grant from Lilly Research Laboratories, Eli Lilly and Company, and in part by a Mental Health Clinical Research Center grant (MH-41115) to the University of Texas Southwestern Medical Center. Appreciation is expressed to Janet H. Potvin, Ph.D., for assistance in the development of the manuscript.

References

- American Psychiatric Association. *DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed., revised. Washington, DC: American Psychiatric Press, 1987.
- Begg, C.B. On inferences from Wei's biased coin design for clinical trials. *Biometrika*, 77:467-484, 1990.
- Coble, P.A.; Kupfer, D.J.; Spiker, D.G.; Neil, J.F.; and McPartland, R.J. EEG sleep in primary depression: A longitudinal placebo study. *Journal of Affective Disorders*, 1:131-138, 1979.

Crowder, M.J., and Hand, D.J. *Analysis of Repeated Measures*. London: Chapman and Hall, 1990.

Dawson, J.D., and Lagakos, S.W. Analyzing laboratory marker changes in AIDS clinical trials. *Journal of AIDS*, 4:667-676, 1991.

Endicott, J., and Spitzer, R.L. A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*, 35:837-844, 1978.

Guy, W., ed. *ECDEU Assessment Manual for Psychopharmacology, Revised*. (DHEW Publication No. [ADM] 76-338) Washington, DC: Superintendent of Documents, U.S. Government Printing Office, 1976. pp. 76-338.

Hamilton, M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 12:278-296, 1960.

Joyce, P.R., and Paykel, E.S. Predictors of drug response in depression. *Archives of General Psychiatry*, 46:89-99, 1989.

Montgomery, S.A., and Åsberg, M.A. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, 134:382-389, 1979.

Press, S.J. *Bayesian Statistics: Principles, Models, and Applications*. New York: John Wiley and Sons, 1989.

Rechtschaffen, A., and Kales, A., eds. *A Manual of Standardized Terminology, Techniques, and Scoring System of Sleep Stages of Human Subjects*. Publication 204. Bethesda, MD: National Institutes of Health, 1968.

Rush, A.J.; Erman, M.K.; Schlessler, M.A.; Roffwarg, H.P.; Nishendu, B.; Manoochehr, K.; Fairchild, C.; and Giles, D.E. Alprazolam vs. amitriptyline in depressions with reduced REM latencies. *Archives of General Psychiatry*, 42:1154-1159, 1985.

Rush, A.J.; Giles, D.E.; Jarrett, R.B.; Feldman-Koffler, F.; Debus, J.R.; Weissenburger, J.; Orsulak, P.J.; and Roffwarg, H.P. Reduced REM latency predicts response to tricyclic medication in depressed outpatients. *Biological Psychiatry*, 26:61-72, 1989.

SAS User's Guide, Statistics, Version 6 Edition. Cary, NC: SAS Institute, Inc., 1990.

Spitzer, R.L., and Endicott, J. *Schedule for Affective Disorders and Schizophrenia-Change*. 3rd ed. New York: New York State Psychiatric Institute, 1978.

Spitzer, R.; Endicott, J.; and Robins, E. *Research Diagnostic Criteria for a Selected Group of Functional Disorders*. 2nd ed. New York: New York State Psychiatric Institute, 1975.

Svendsen, K., and Christensen, P.G. Duration of REM sleep latency as predictor of effect of antidepressant therapy: A preliminary report. *Acta Psychiatrica Scandinavica*, 64:238-243, 1981.

U.S. Food and Drug Administration. *COSTART Coding Symbols for Thesaurus of Adverse Reaction Terms*. 2nd ed. Rockville, MD: Food and Drug Administration, 1985.

Ware, J., and Sherbourne, C.D. The MOS 36-item short-form health survey (SF-36). *Medical Care*, 30:473-483, 1992.

Ware, J. Investigating therapies of potentially great benefit: ECMO. *Statistics in Science*, 4:298-340, 1989.

Wei, L.J., and Durham, S. The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, 73:840-843, 1978.

Wolfram, S. *Mathematica™: A System for Doing Mathematics by Computer*. Redwood City, CA: Addison-Wesley Publishing Co., Inc., 1988.

Wu, M.C., and Bailey, K. Analyzing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7:337-346, 1988.