Acute Worsening of Chronic Depression During a Double-Blind, Randomized Clinical Trial of Antidepressant Efficacy: Differences by Sex and Menopausal Status

Anne T. Harvey, Ph.D.; Beryl S. Silkey, Sc.M.;
Susan G. Kornstein, M.D.; and Cathryn M. Clary, M.D.

Objective: Acute worsening of depression can negatively impact the outcomes of clinical trials of antidepressants and patient compliance to treatment. We hypothesized that acute worsenings would be more frequent in premenopausal women, relative to men or postmenopausal women, and in women who had demonstrated premenstrual symptom exacerbations (PMEs) prior to treatment, relative to those who had demonstrated no PMEs.

Method: Subjects diagnosed with DSM-III-R chronic major depressive disorder or *double depression* (dysthymia with concurrent major depressive episode) were randomly assigned between February 1993 and December 1994 to 12 weeks of double-blind treatment with flexibly-dosed sertraline or imipramine, with crossover to the alternate drug in the absence of response. A 6-point or more increase in the 17-item Hamilton Rating Scale for Depression relative to the (7–14 day) previous visit defined worsening. PME was assessed through daily diaries prior to treatment.

Results: There were 3582 evaluable visits attended by 554 subjects. Premenopausal women had a deteriorating depressive presentation at a greater proportion of their visits (8.6%) than did postmenopausal women (4.5%, $p < .01$) or men (5.9%, $p < .01$). The presence of PME at baseline was associated with more worsenings than the absence of PME (12.0% vs. 7.3%, $p < .05$). Results were similar whether the subject was treated with sertraline or imipramine. Nonresponse at treatment completion was more likely among subjects with worsening ($p < .01$). Dropouts were more likely than completers to have had an exacerbation at their terminal visit ($p < .05$).

Conclusion: Acute worsening of depression was associated with reproductive variables and negatively affected clinical trial outcomes including early treatment discontinuation and nonresponse.

(J Clin Psychiatry 2007;68:951–958)

Received April 11, 2006; accepted Nov. 22, 2006. From Via Christi Research, Inc., Wichita, Kan. (Dr. Harvey and Ms. Silkey); the Department of Psychiatry, Virginia Commonwealth University, Richmond, Va. (Dr. Kornstein); and Pfizer Inc., New York, N.Y. (Dr. Clary).

This research was supported by Pfizer Inc., New York, N.Y.

A preliminary report of this study was presented in poster format at the 45th annual meeting of the New Clinical Drug Evaluation Unit, June 6–9, 2005; Boca Raton, Fla.

Dr. Kornstein has received grant/research support from Pfizer, Bristol-Myers Squibb, Eli Lilly, Forest Laboratories, GlaxoSmithKline, Mitsubishi-Tokyo, Pharmacia and Upjohn, Merck, Biovail Laboratories, Wyeth, Berlex Laboratories, Novartis, Sepracor, Boehringer-Ingelheim, Sanofi-Synthelabo, and AstraZeneca and has been a member of advisory boards for Pfizer, Wyeth, Eli Lilly, Bristol-Myers Squibb, Warner-Chilcott, Biovail Laboratories, Berlex Laboratories, and Forest Laboratories. Dr. Clary is a stock shareholder of Pfizer, and her spouse is a consultant for Pfizer. Dr. Harvey and Ms. Silkey report no additional financial or other relationships relevant to the subject of this article.

Corresponding author and reprints: Anne T. Harvey, Ph.D., Via Christi Research, Inc., 1100 N. St. Francis, Suite 200, Wichita, KS 67214 (e-mail: annie_harvey@via-christi.org).

Sudden worsening of a depressive syndrome can compromise the efficient conduct of clinical trials of antidepressant efficacy. Worsening may discourage a subject and lead to treatment noncompliance and/or premature withdrawal from treatment trials. Because efficacy is typically based on the change from baseline of a single measure of depressive severity at the end of treatment, worsening at the terminal visit can convert a responder (≥ 50% reduction in Hamilton Rating Scale for Depression[1] score) to a nonresponder. In some clinical trials—and in clinical practice[2]—deterioration could lead to a change in dose or a change in treatment. Finally, the power to separate the signal of an efficacious drug from that of placebo is weakened by the resulting increased variability of response.[3]

Exacerbation during short-term treatment trials of antidepressants has been examined from several perspectives. Worsening during a drug-free interval between screening and baseline has been linked to poor separation of subsequent drug response from placebo response.[4] Worsening following an initial response (*nonpersistent response*) has been used to exclude subjects from classification as *true drug* responders.[5] Worsening in the first 6 weeks of treatment predicted nonresponse to 12 weeks of fluoxetine and a greater likelihood of dropout prior to study completion.[6] In a preliminary study comparing potential correlates of

worsening, reproductive status had a greater influence than did concurrent life events or a change in the clinician-rater[7]: worsening was more common in women with natural menstrual cycles than in noncycling women or men, and showed a large effect of menstrual cycle phase.

Here we examine data from a study comparing the efficacy of sertraline and imipramine in subjects with chronic depression. A subset of the premenopausal women had maintained daily logs at baseline that allowed prospective identification of premenstrual symptom exacerbations (PMEs) (Kornstein et al.[8]). We hypothesized that among subjects with chronic depression, acute worsenings would be more frequent in premenopausal women than in men or postmenopausal women. Among premenopausal women, we hypothesized that acute worsenings would be more frequent in those who had demonstrated PMEs prior to treatment, relative to those who had demonstrated no PMEs.

METHOD

Clinical Trial

Following institutional review board approval, individuals aged 21 to 65 years who met DSM-III-R criteria for chronic major depressive disorder (MDD) (i.e., major depressive episode of at least 2 years' duration without antecedent dysthymia) or double depression (i.e., major depressive episode superimposed on dysthymia) in the absence of significant medical or psychiatric comorbidity were recruited through advertisement or medical referral to 1 of 10 university-affiliated medical centers or 2 clinical research centers. All subjects provided written informed consent. The rationale, design, methods, and inclusion/exclusion criteria of the treatment trial have been reported in detail elsewhere.[9,10]

A 24-item Hamilton Rating Scale for Depression ($HAM\text{-}D_{24}$)[1] score of 18 or more and a Clinical Global Impressions-Severity of Illness scale (CGI-S)[11] score of 3 or more after 1 week of placebo washout were required for study entry. Between February 1993 and December 1994, subjects were randomly assigned to 12 weeks of double-blind treatment with sertraline or imipramine in a 2:1 ratio (acute phase). Nonresponders after 12 weeks were removed from study medication for at least a week and then were eligible to enter a 12-week double-blind trial of the alternate medication (crossover phase).

The starting dose of both medications was 50 mg/day. Dose titration was allowed according to clinical response and the absence of dose-limiting side effects, to a maximum of 300 mg/day imipramine or 200 mg/day sertraline. The mean (SD) sertraline dose at the end of the initial 12-week trial was 140 (60) mg/day in women and 143 (59) mg/day in men. The mean (SD) imipramine dose at the end of the initial 12-week trial was 196 (82) mg/day in women and 208 (83) mg/day in men.

Subject visits were scheduled at weekly intervals for the first 6 weeks and biweekly thereafter, in both the acute and (if needed) the crossover phases. At baseline and the week 1, 2, 4, 6, 8, 10, and 12 visits of each phase, rating scales included the $HAM\text{-}D_{24}$ and the CGI-S and (post-baseline) the Clinical Global Impressions-Improvement (CGI-I)[11] scales. Subjects were recruited to the crossover phase in the absence of a satisfactory therapeutic response, defined as a decrease in the $HAM\text{-}D_{24}$ score of ≥ 50%, a $HAM\text{-}D_{24}$ score of 15 or more, a CGI-S score of 3 or more, and a CGI-I score of 1 or 2 (corresponding to "very much" or "much improved") at 2 consecutive visits.

Assignment of Reproductive Status

For purposes of this analysis, subjects were classified as male, premenopausal, or postmenopausal. A menstrual history obtained at the screening visit provided information regarding menopausal and hysterectomy status, date of last menses, and cycle length. Irregularly menstruating women were excluded from classification by requiring evidence of regular menstrual cycles (menses within the 35 days previous to the screening visit, reported cycle length, and/or examination of the daily symptom logs described below). Inclusion as postmenopausal required either endorsement of postmenopausal state with last menses at least 1 year previous or reported hysterectomy in women 52 years or older.

Assessment of Premenstrual Symptom Exacerbation at Baseline

A Daily Log of Mood Symptoms was mailed to consenting study participants (men and women) at the time their screen visit was scheduled. Logs containing a premenstrual and midfollicular interval prior to week 2 on study drug were used to define PME status at baseline (only 10% of premenopausal women had achieved a satisfactory therapeutic response by that point). The logs allowed prospective identification of a premenstrual exacerbation of selected symptoms.[8] All subjects were asked to rate daily the symptoms depressed/sad mood, anxiety, irritability, mood swings, and fatigue. Menstruating woman were further asked to indicate days on which bleeding occurred and to rate physical symptoms experienced and related to the menstrual cycle. Each symptom was rated from 0 (none) to 5 (extreme). Subjects were informed that the log "helps us understand the fluctuation in your symptoms from day to day and week to week" and were not made aware of the interest in the menstrual cycle.[12] Premenopausal nonusers of oral contraceptives were assessed for PME, defined by (1) a change of 10 points in total daily log score from follicular phase (day 6 through day 10 of the menstrual cycle, with the onset of bleeding marking day 1) to luteal phase (day –5 through day –1), and (2) severity of 4 or 5, on at least 2 of 5 luteal days, in at least 1 symptom other than depressed mood.

Assessment of Worsening of MDD

Each postbaseline visit in the acute and crossover phases was assessed for the presence of a clinically significant worsening of MDD. Worsening was defined as a 6-point or more increase in HAM-D_{17} relative to the (7–14 day) previous visit. Worsening was defined in terms of the HAM-D_{17} because the HAM-D_{17} is the primary outcome measure in clinical trials of antidepressant efficacy. Clinical significance of the worsening was addressed by requiring a HAM-D_{17} change of magnitude previously associated with a 1-point increase in the CGI-S: a 6-point or more increase in HAM-D_{17}.[7] A criterion of 7 to 14 days between eligible visits was chosen on the basis of the 1- to 2-week visit intervals typical of acute treatment trials and on the sensitivity of this interval to menstrually related exacerbations.

Statistical Analysis

Worsening was examined in terms of reproductive state (premenopausal, postmenopausal, male) and whether PME had been demonstrated at baseline. The visit was chosen as the unit of study for the primary analysis because our interest was in sudden worsening (at an individual visit as opposed to averaged over a subject's multiple visits) as an impetus for treatment adherence and response assessment. The primary analysis was repeated using the subject as the unit of analysis. Continuous variables were compared using analysis of variance for parametric and Kruskal-Wallis tests for nonparametric variables, and Tukey's honestly significant difference (HSD) for post hoc testing. Categorical variables were compared using χ^2 tests. To examine the effect of reproductive status (male, premenopausal, postmenopausal) on exacerbations after controlling for the effects of age, we used a repeated-measures general linear model (generalized estimating equations) for binary variables. This analytic method addresses the correlation of repeated measurement on the same individual over time and handles differing numbers of repeated measures per individual. The evaluations of exacerbation (yes/no) at each visit were considered the dependent variable. The model included reproductive status as the independent variable, and age as a covariate. Pairwise comparisons of the reproductive groups were made using Wald χ^2 tests. Effect sizes were calculated following the methods of Cohen.[13] The level of statistical significance was set to $p < .05$ throughout. SPSS software, version 15.0 (SPSS Inc., Chicago, Ill.) was used for all analyses.

RESULTS

Classification of Subjects by Reproductive Category

Of the 635 subjects enrolled in the study, 90% could be categorized as male (N = 235), premenopausal (N = 275) or postmenopausal (N = 61). The women who could not be classified (1) were under 52 years old, had endorsed hysterectomy, and had unknown ovarian status (N = 36); (2) had incomplete or inconsistent reproductive data (N = 20); or (3) were menstruating but last menses was more than 35 days prior to the screening visit and they were deemed irregularly menstruating (N = 8).

Retention of Subjects and Visits

The 571 reproductively classified subjects attended 5271 visits from acute baseline through completion of the acute and crossover phases. Across reproductive categories there was no difference in the total number of visits and no difference in retention at various critical points in the study. When visits were limited to those that (1) occurred following at least 1 week of treatment, (2) followed a previous visit by 7 to 14 days, and (3) had HAM-D_{17} ratings available at both visits, 3582 evaluable visits remained, contributed by 554 subjects.

Comparison of Reproductive Groups on Baseline and Demographic Characteristics

Baseline differences between men and women have been detailed in a previous article.[14] When menopausal status was also taken into account, each reproductive group differed in mean age from every other group (Table 1). All remaining baseline comparisons with statistically significant overall differences were further examined using age as a grouping variable. We chose to define age group intervals on the basis of (1) existing cut points in the data defining premenopausal and postmenopausal women, and (2) inclusion of at least 10% of the reproductive group in each relevant interval. Final age group intervals were 21–29, 30–35, 36–40, 41–45, 46–52, and 53–65 years.

The sample was overwhelmingly white, and the majority of subjects were employed. Reproductive categories were similar in terms of number of previous depressive episodes, length of the current episode, prior psychiatric hospitalization, presence of any Axis II pathology, depressive subtype, and depressive severity.

Men were more often married than were premenopausal women in the 36–40 and 46–52 year age groups (42% vs. 27%, $p < .05$ and 53% vs. 25%, $p < .05$, respectively) and more often than postmenopausal women in the 53–65 year age group (51% vs. 33%, $p < .05$). Within all age groups, women reported an earlier age at onset of depression (mean = 2.5 years) and premenopausal women an earlier age at onset of dysthymia (mean = 3.8 years). A history of alcohol abuse or dependence was most common among men within each age group. Within all age groups a higher proportion of women had a family history of affective disorder.

Worsening of Depression and the Effect of Reproductive Status

Worsening of depression as indicated by a 6-point or more increase in HAM-D_{17} was scored at 7.1% of evalu-

Table 1. Sociodemographic and Clinical Characteristics of Subjects, by Reproductive Category (N = 554)

Characteristic	Male (N = 229)	Postmenopausal (N = 58)	Premenopausal (N = 267)	Statistical Test	p Value
Sociodemographic variables[a]					
Age, y	43.4 (9.7)	55.6 (4.4)	36.8 (7.6)	KW	< .001[b]
White, %	90.4	91.4	92.5	χ^2	.70
Married, %	44.5	36.2	30.3	χ^2	< .01[c]
Employed in occupation, %	58.1	58.6	59.6	χ^2	.95
Psychiatric history[a]					
Age at onset of depression, y	27.3 (12.5)	32.4 (15.4)	21.1 (9.5)	KW	< .001[d]
Age at onset of dysthymia, y	19.0 (13.5)	27.8 (18.4)	13.0 (9.9)	KW	< .001[e]
No. of previous episodes of depression	1.7 (2.2)	2.0 (2.1)	1.8 (2.1)	KW	.24
Duration of current episode, y	5.9 (8.1)	6.6 (9.0)	5.6 (7.4)	KW	.99
Previous psychiatric hospitalization, %	10.9	5.2	8.2	χ^2	.33
Axis II disorder present, %	44.1	32.8	43.8	χ^2	.26
History of alcohol abuse/dependence, %	38.4	10.3	25.8	χ^2	< .001[f]
History of drug abuse/dependence, %	21.4	8.6	16.5	χ^2	.06
First-degree relative with affective disorder, %	51.8	57.4	67.2	χ^2	< .01[g]
Depressive measures at baseline[a]					
Double depression, %	59.4	46.6	51.3	χ^2	.09
HAM-D_{24} score	24.8 (5.2)	24.7 (5.0)	25.5 (5.0)	ANOVA	.29
HAM-D_{17} score	18.5 (4.1)	18.7 (3.9)	18.7 (3.8)	ANOVA	.83
No. of evaluable visits	6.4 (3.0)	6.5 (3.5)	6.6 (3.1)	ANOVA	.77

[a]Data expressed as mean (SD) unless otherwise specified.
[b]Each group differs from every other.
[c]When examined within age group, men were more frequently married than premenopausal females in the 36–40 and 46–52 year age groups, and more often than postmenopausal women in the 53–65 year age group.
[d]Within all age groups, females reported an age at onset a mean of 2.5 years earlier than men.
[e]Within all age groups, premenopausal females reported an age at onset a mean of 3.8 years earlier than men.
[f]Within all age groups, women reported a lower prevalence.
[g]Within all age groups, a higher proportion of women had a family history of affective disorder.
Abbreviations: ANOVA = analysis of variance, HAM-D_{17} = 17-item Hamilton Rating Scale for Depression, HAM-D_{24} = 24-item Hamilton Rating Scale for Depression, KW = Kruskal-Wallis test.

Table 2. Worsening of Depression at Postbaseline Visits, by Reproductive Category (N = 3582)

				Premenopausal Subcategories		
Variable	Male	Postmenopausal	Premenopausal	With Premenstrual Symptoms	Without Premenstrual Symptoms	Premenstrual Symptoms Not Assessed
Reproductively classified subjects, N	235	61	275	30	67	178
Subjects with postbaseline visits, N	231	59	272	30	67	223
Postbaseline visits, N[a]	1836	469	2247	297	626	1324
Evaluable postbaseline visits, N[b]	1456	374	1752	233	507	1012
Visits with deterioration, N[c]	86	17	150	28	37	85
Visits with deterioration, %	5.9[d]	4.5[e]	8.6	12.0[f]	7.3	8.4

[a]Excludes acute baseline and crossover baseline.
[b]Evaluable visits occurred between 7 and 14 days from the previous visit; both visits captured HAM-D_{17}.
[c]Increase of 6 points or more in the HAM-D_{17} from previous visit.
[d]Premenopausal > male: 2-tailed p = .004, effect size = 0.10.
[e]Premenopausal > postmenopausal: 2-tailed p = .008, effect size = 0.17.
[f]With premenstrual symptoms > without premenstrual symptoms: 2-tailed p = .049, effect size = 0.16.
Abbreviation: HAM-D_{17} = 17-item Hamilton Rating Scale for Depression.

able visits. As shown in Table 2, depressed premenopausal women presented this deteriorating clinical picture at 8.6% of their postbaseline visits. This was higher than the rate of deterioration experienced by postmenopausal women (4.5%, p < .01) or by men (5.9%, p < .01). Among women charting symptoms across the menstrual cycle at baseline, worsenings were more likely at the postbaseline visits of those with documented PME (12.0% vs. 7.3%, p < .05).

When examined by age group as shown in Figure 1, the rate of worsening was consistently highest in premenopausal women and decreased with age. The effect of reproductive status on worsenings was subsequently assessed while controlling for age, with the overall effect of premenopausal status achieving marginal significance (p = .07). The difference between premenopausal and postmenopausal women continued to be statistically significant (p = .03) after adjustment for age.

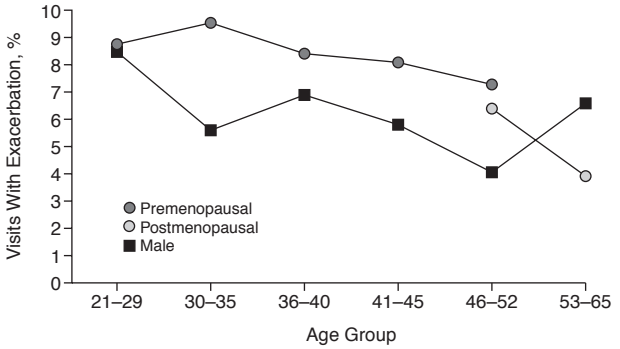
Worsening of Depression Using the Subject as the Unit of Analysis

Mirroring the results when using visits as the unit of analysis but with the lower power of a subject-based

analysis (i.e., fewer data points), a greater proportion of premenopausal women experienced at least 1 exacerbation when compared to men (40.8% vs. 32.3%, p = .06) or to postmenopausal women (40.8% vs. 24.1%, p = .02); women with PME at baseline were numerically but not statistically more likely to have an exacerbation (53.3% vs. 41.8%, p = .38).

The potential for concentration of exacerbations within some subjects was further examined by identifying subjects who experienced more than 1 exacerbation. No subjects with 0 to 4 evaluable visits experienced multiple exacerbations (Table 3). In subjects with 5 to 9 evaluable visits, a greater proportion of premenopausal women had multiple exacerbations, a trend that reached statistical significance among subjects with 10 to 16 evaluable visits. Thus, the preponderance in premenopausal women of visits with exacerbations was influenced both by a greater number of women experiencing exacerbations, and by a greater number of exacerbations experienced by individual women.

Figure 1. Visits at Which Depressive Exacerbation Was Evident as a Function of Age, by Reproductive Status, %[a]



[a]Exacerbation was defined as an increase in HAM-D_{17} of 6 points or more from a visit 7–14 days previous.
Abbreviation: HAM-D_{17} = 17-item Hamilton Rating Scale for Depression.

Effect of Sertraline or Imipramine Treatment on Worsening

The rank frequency of visits with worsening was the same regardless of whether the subject was receiving treatment with sertraline or imipramine (premenopausal > male > postmenopausal, and PME at baseline > no PME at baseline) (Figure 2). At visits on sertraline, only the contrast between women with and without PME at baseline was statistically significant (12.3% vs. 6.4%, p < .05). At visits on imipramine, only the contrast between premenopausal women and men was statistically significant (9.3% vs. 5.9%, p < .05).

Association of Worsening With Response Status

Completers and noncompleters of drug treatment were examined separately to assess the relationship between worsening and response, defined as a decline in HAM-D_{17} score from baseline of ≥ 50%. In subjects who completed the acute phase (12 weeks' treatment with sertraline or imipramine), nonresponse was 1.5 times more likely among those with worsening(s) (p < .001; Table 4). Subjects who also completed the crossover phase (12 weeks' treatment with each drug) had a 1.6 greater probability of nonresponse in the presence of 1 or more exacerbations (p = .005). Few subjects who dropped out during the acute or crossover phase were responders; the relative risk of nonresponse in subjects with exacerbation(s) was 1.1 (not significant).

Worsening at the Terminal Visit

Terminal visits met evaluability criteria for significantly more completers than dropouts (76% vs. 58%, p < .001). Completers were more likely to experience an exacerbation at a nonterminal visit than at a terminal visit (p = .02, Table 5). Of the 12 completers who did experience exacerbations at the terminal visit, 8 were nonresponders and 5 of the 8 would have been considered responders at their previous visit.

Dropouts had exacerbations more frequently at terminal than at nonterminal visits, though the difference was not statistically significant (p = .66). Dropouts were more likely than completers to have had an exacerbation at their terminal visit (p = .04).

DISCUSSION

In this clinical trial comparing sertraline and imipramine in the treatment of chronic depression, subjects presented with a clinically significant acute worsening at 7.1% of postbaseline visits. As hypothesized, acute worsening was more frequent in premenopausal women (8.6% of visits) than in postmenopausal women or men (4.5% and 5.9% of visits, respectively). PME at baseline further increased the risk of worsening of the depressive syndrome (to 12% of visits) as measured by the HAM-D_{17}. Differences across reproductive groups in the rate of worsening were the same whether treatment was with sertraline or imipramine. Results were similar whether the unit of study was the subject or the visit. These results confirm a preliminary report of an excess of acute worsenings of MDD among premenopausal clinical trial participants with MDD.[7] The effect size of having PME at baseline was sufficient to explain the rate differences between premenopausal women and menopausal women and men, suggesting that any premenstrual changes that continued through the study could be sufficient to explain the higher rate of exacerbation in premenopausal women.

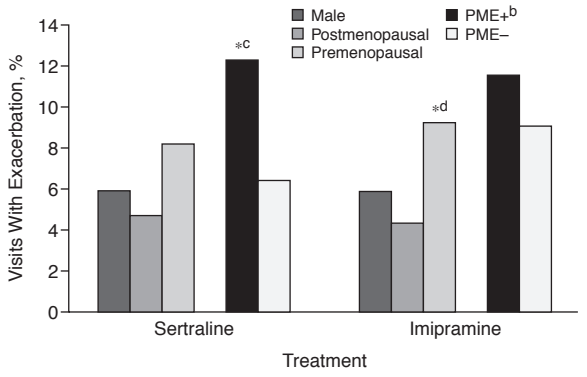
The relevance of worsening in this study was examined in relation to response to treatment, early termina-

Table 3. Presence of Multiple Exacerbations, by Reproductive Category[a]

No. of evaluable visits[b]	Subjects With Multiple Exacerbations, n/N (%)[c] Male	Postmenopausal	Premenopausal	p Value
0–4	0/58 (0)	0/18 (0)	0/63 (0)	1.0
5–9	3/134 (2)	0/29 (0)	10/152 (7)	.13
10–16	7/37 (19)	2/11 (18)	22/52 (42)	.048

[a]Subjects who responded to acute treatment required no more than 7 evaluations.
[b]Evaluable visits occurred between 7 and 14 days from the previous visit; both visits captured HAM-D_{17}.
[c]Increase of 6 points or more in HAM-D_{17} from previous visit.
Abbreviation: HAM-D_{17} = 17-item Hamilton Rating Scale for Depression.

Figure 2. Visits at Which Exacerbation Was Evident as a Function of Treatment With Sertraline (2673 visits of 422 subjects) or Imipramine (1649 visits of 277 subjects), by Reproductive Status, %[a]



[a]Exacerbation was defined as an increase in HAM-D_{17} of 6 points or more from a visit 7–14 days previous.
[b]A subset of premenopausal women (85 taking sertraline and 48 taking imipramine) maintained daily logs of premenstrual symptoms at baseline that allowed their classification as having (PME+, 233 visits of 43 women) or not having (PME–, 507 visits of 90 women) premenstrual symptoms.
[c]In the sertraline treatment group, PME+ differed from PME–.
[d]In the imipramine treatment group, premenopausal women differed from men.
*$p < .05$.
Abbreviations: HAM-D_{17} = 17-item Hamilton Rating Scale for Depression, PME+ = women with premenstrual symptoms, PME– = women without premenstrual symptoms.

Table 4. Response Rate of Subjects With and Without Worsening

Category	Worsening(s), n/N (%) Yes	No	p Value
Completers			
Acute phase[a]	65/155 (42)	182/296 (61)	< .001
Crossover phase	9/35 (26)	45/82 (55)	< .01
Dropouts[b]	2/38 (5)	13/96 (14)	NS

[a]Includes subjects who continued to crossover phase.
[b]Discontinued in acute or crossover phase.
Abbreviation: NS = not significant.

Table 5. Worsening at Terminal and Nonterminal Visits of Completers and Dropouts

Category	Terminal Visit, n/N (%)	Nonterminal Visit, n/N (%)	p Value
Completers	12/318 (3.8)	193/2705 (7.1)	.02
Dropouts	8/78 (10.3)[a]	40/481 (8.3)	.66

[a]$p = .04$ vs. terminal visits of completers.

tion, and nondrug correlates of worsening, each of which was also examined by Cusin et al.[6] during a 12-week treatment trial with fluoxetine. In that study, response was achieved in 75% of subjects without, but only 58% of subjects with worsening during the first 6 weeks of treatment. The current study supports that finding: subjects completing 12 weeks of treatment had a 62% response rate in the absence but only a 42% response rate in the presence of worsening at any eligible visit over the 12 weeks. In our study only 12 subjects (4%) completing the prescribed course of treatment had exacerbations at the terminal visit; hence, the difference in response rate cannot be attributed solely to worsening at the terminal visit. Eight of the 12 completed as nonresponders. Notably, 5 of the 8 had merited responder status at the penultimate visit thus demonstrating that worsening at the terminal visit can convert responders to nonresponders.

We found a higher rate of worsening at the terminal visits of dropouts when compared to completers (10.3% vs. 3.8%) consistent with worsening leading to early termination. This comparison probably understates the association by not reflecting dropouts who worsened between scheduled visits and were lost to follow-up. In addition, a large proportion of the dropouts lacked evaluable terminal visit information. Missing data could explain the discrepancy found when the dropout rates of the earlier fluoxetine study[6] (37% of subjects without, 50% of subjects with early worsening) were compared with the dropout rates in the acute phase of the current study (22% of subjects without, and 12% of subjects with any worsening).

In this study, premenopausal status was associated with the presence of worsening, even after adjustment for age. The earlier study with fluoxetine[6] reported instead a predominance of men among subjects with exacerbations, and no association between worsening and age. This difference may be accounted for by their exclusion from

analysis of the 17% of subjects who experienced exacerbations at week 1, or beyond the 6th week of treatment; the high proportion of men (68%) in the remaining sample suggests that that many women with exacerbations were excluded.

One explanation for the high rate of worsening experienced by premenopausal women in the current study is the presence of change associated with the menstrual cycle. This notion is supported by the finding that women who had prospectively demonstrated PME at baseline showed the highest rate of HAM-D_{17} worsening during treatment, consistent with continuation of cyclic symptoms. Further, the rate of worsening differed between premenopausal and postmenopausal women even after controlling for age. Unfortunately, the data precluded direct assessment of the contribution of perimenstrual increases in HAM-D_{17} to worsening of depression, which was assessed without reference to the menstrual cycle. In some respects ascribing worsening to the menstrual cycle is at odds with the literature and with other analyses of the PME data. Studies of MDD concurrent with premenstrual symptoms conclude that effective treatment of underlying depression is associated with diminution of premenstrual symptoms.[15–17] Indeed, a separate analysis of the current study focusing on the measure of PME rather than on the HAM-D_{17} also confirmed a diminution, finding that response to antidepressant treatment was associated with an 80% decrease in the rate of PME measured at response endpoint (Kornstein et al.[8]). The persistence of HAM-D_{17} worsening in women with PME at baseline could represent a nonspecific tendency to a fluctuating course of illness,[2,18,19] although post hoc analyses revealed no relationship between the presence of exacerbations and the Structured Clinical Interview for DSM-III-R Personality Disorders[20] diagnostic measure of borderline personality. Alternatively, some women are reported to retain premenstrual symptoms in spite of effective antidepressant treatment,[17,21] an observation that may reflect the presence of 2 separate illnesses.

Because of their efficacy in premenstrual syndromes, serotonin reuptake inhibitors have been believed to reduce premenstrual worsening of depression. By this line of reasoning sertraline, relative to imipramine, would be expected to reduce cyclic differences between women with and without PME at baseline, a position supported by a small study using the methodology described here.[22] Such was not the case. The difference between PME groups in the incidence of worsening was greater during sertraline treatment (PME: 12.3%, no PME 6.4%, $p < .05$) than during imipramine (PME: 11.6%, no PME 9.1%, $p = .53$). It may be that PME at baseline signaled a general instability of depressive course. In addition, there may have been some contribution of cyclic pharmacokinetic[23] changes to drug response, an issue that has not been explored with either drug.

The primary limitation of this study was the lack of information that might have illuminated the potential causes of worsening. It was not possible to place visits within the context of the menstrual cycle. Changes in drug concentrations attendant to the flexible dose design or to noncompliance were not addressed in the current analysis. Other unexplored factors that might have led to acute changes in depressive severity include life events, a change in HAM-D rater, and undiagnosed bipolar illness.[24] Generalization was limited by exclusion of nearly 20% of the women, many of whom were probably perimenopausal.

The methodology of the current study holds promise for the assessment of perimenstrual worsening of depression during clinical trials, requiring only that dates of menses be routinely collected with other vital signs. Assessment of PME has been based to date on methods developed for premenstrual syndromes in PMS samples that routinely exclude depressed women. A major limitation of the PMS-based methods is the requirement for daily charting over 1 or more cycles, which is logistically difficult and imposes a self-selection bias in the sample. Further, the symptoms typically evaluated are those associated with PMS and not necessarily with depression. Finally, the specific intervals chosen to define the change in symptom severity may not be valid when depression as measured by the HAM-D_{17} is the construct of interest. The approach used in the current analysis would largely circumvent these limitations.

Our data suggest that premenstrual change may be one source of worsened depression in young women. Though consensus guidelines for the treatment of premenstrual worsening have been published,[25] there are no evidence-based guidelines. Nonetheless, it would appear that when a woman of reproductive age presents with an apparent setback in her depressive syndrome, it could be helpful to determine when her last menses occurred. If it has been several weeks her setback may be of a transient nature and handled with watchful waiting, with temporary dose increases or with temporary adjunctive medication as needed. Women should be counseled to consider the timing and potential transience of their symptom worsening and encouraged to participate in their symptom tracking and treatment decisions. This patient involvement could reduce treatment discontinuation triggered by an unexpected worsening of symptoms. If a variable dose schedule presents a problem, permanent dose increases are an alternative but can increase cost and exposure to medication that may not be needed for most of the month. Switching to a new antidepressant premenstrually should be discouraged; rather, reassessment should take place just after menses. Patient reporting of negative life events (e.g., discussions about divorce) may increase prior to menses; these issues should be revisited at a less symptomatic stage of the cycle.

In conclusion, we found that worsening of the depressive syndrome during drug treatment occurred most frequently in premenopausal women. This was particularly so in those with premenstrual symptom exacerbations, notwithstanding treatment with a drug with proven efficacy in PMS. Worsening was associated with clinical outcomes including early treatment discontinuation and nonresponse. Attention to dates of menses and other potential sources of acute worsening will lead to a better understanding of the short-term course of depression and allow development of rational measures to enhance the precision of clinical trials.

Drug names: fluoxetine (Prozac and others), imipramine (Tofranil and others), sertraline (Zoloft and others).

REFERENCES

1. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
2. Benedetti F, Barbini B, Colombo C, et al. Infradian mood fluctuations during a major depressive episode. J Affect Disord 1996;41:81–87
3. Kraemer HC, Thiemann S. A strategy to use soft data effectively in randomized controlled clinical trials. J Consult Clin Psychol 1989;57:148–154
4. Evans KR, Sills T, Wunderlich GR, et al. Worsening of depressive symptoms prior to randomization in clinical trials: a possible screen for placebo responders? J Psychiatr Res 2004;38:437–444
5. Quitkin FM, Rabkin JD, Markowitz JM, et al. Use of pattern analysis to identify true drug response: a replication. Arch Gen Psychiatry 1987;44:259–264
6. Cusin C, Perlis RH, Amsterdam JD, et al. Early symptomatic worsening during treatment with fluoxetine in major depressive disorder: prevalence and implications [poster]. Presented at the 45th annual meeting of the New Clinical Drug Evaluation Unit; June 6–9, 2005; Boca Raton, Fla
7. Harvey AT. Contribution of the menstrual cycle to instability of depressive severity [poster]. Presented at the 39th annual meeting of the New Clinical Drug Evaluation Unit; June 1999; Boca Raton, Fla
8. Kornstein SG, Yonkers KA, Schatzberg AF, et al. Premenstrual exacerbation of depression. Presented at the 148th Annual Meeting of the American Psychiatric Association; May 22, 1995; Miami, Fla. Abstract No. 87C:147
9. Rush AJ, Koran LM, Keller MB, et al. The treatment of chronic depression, pt 1: study design and rationale for evaluating the comparative efficacy of sertraline and imipramine as acute, crossover, continuation, and maintenance phase therapies. J Clin Psychiatry 1998;59:589–597
10. Keller MB, Gelenberg AJ, Hirschfeld RMA, et al. The treatment of chronic depression, pt 2: a double-blind, randomized trial of sertraline and imipramine. J Clin Psychiatry 1998;59:598–607
11. Guy W. ECDEU Assessment Manual for Psychopharmacology. US Department of Health, Education, and Welfare publication. (ADM) 76-338. Rockville, Md: National Institute of Mental Health; 1976:218–222
12. Ruble DN. Premenstrual symptoms: a reinterpretation. Science 1977;197:291–292
13. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988
14. Kornstein SG, Schatzberg AF, Thase ME, et al. Gender differences in chronic major and double depression. J Affect Disord 2000;60:1–11
15. Freeman EW, Rickels K, Sondheimer SJ, et al. Nefazodone in the treatment of premenstrual syndrome: a preliminary study. J Clin Psychopharmacol 1994;14:180–186
16. Lopez-Mato A, Vieitez A, Marquez C. Differential response of fluoxetine in premenstrual dysphoric disorder (PMDD) and catamenial exacerbation of depression. Arch Womens Men Health 2001;3(suppl 2):67
17. Glick R, Harrison W, Endicott J, et al. Treatment of premenstrual dysphoric symptoms in depressed women. J Am Med Womens Assoc 1991;46:182–185
18. Huba GJ, Lawlor WG, Stallone F, et al. The use of autocorrelation analysis in the longitudinal study of mood patterns in depressed patients. Br J Psychiatry 1976;128:146–155
19. Eastwood MR, Whitton JL, Kramer PM, et al. Infradian rhythms. Arch Gen Psychiatry 1985;42:295–299
20. Spitzer RL, Williams JBW, Gibbon M. Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). New York, NY: Biometrics Research, New York State Psychiatric Institute; 1986
21. Yonkers KA, White K. Premenstrual exacerbation of depression: one process or two? J Clin Psychiatry 1992;53:289–292
22. Harvey AT. Preliminary data suggest efficacy of serotonin reuptake inhibitors in perimenstrual exacerbation of depression [poster]. Presented at the Second World Congress of Women's Mental Health; March 2004; Washington, DC
23. Kimmel S, Gonsalves L, Youngs D, et al. Fluctuating levels of antidepressants premenstrually. J Psychosom Obstet Gynaecol 1992;13:277–280
24. Akiskal HS, Maser JD, Zeller PJ, et al. Switching from "unipolar" to bipolar II: an 11-year prospective study of clinical and temperamental predictors in 550 patients. Arch Gen Psychiatry 1995;52:114–123
25. Steiner M, Pearlstein T, Cohen LS, et al. Expert guidelines for the treatment of severe PMS, PMDD and comorbidities: the role of SSRIs. J Womens Health (Larchmt) 2006;15:57–69

Editor's Note: We encourage authors to submit papers for consideration as a part of our Focus on Women's Mental Health section. Please contact Marlene Freeman, M.D., at marlenef@email.arizona.edu.

