CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-235

STATISTICAL REVIEW(S)

11/09/00

COMPLETED NOV 17 2000

Statistical Review and Evaluation

NOV 17 2000

NDA#	21-235
Submission Date	Mar 20, 2000
Due Date	Jan 14, 2001
Sponsor	Eli Lilly and Company
Name of Drug	Prozac (Fluoxetine Hydrochloride)
Indication	Treatment of major depression
Documents Reviewed	The findings from the statistical analyses.

Introduction

Results of one multicenter, double-blinded, randomized, parallel clinical study (Study B1Y-MC-HCIZ) consisted of three treatment groups (Placebo, Fluoxetine 20-mg daily, Fluoxetine 90-mg once weekly) were submitted to demonstrate the effectiveness of a weekly 90-mg oral dose of enteric-coated fluoxetine hydrochoride for continuation treatment of major depression disorder in patients identified as having responded to acute treatment with oral daily 20-mg doses of marketed formulation of fluoxetine hydrochloride. The study was conducted in 42 U.S. study centers.

The study consisted of four phases. Phase 1 was a 4- to 14- day screening phase. Phase 2 was a 13-week acute treatment phase during which all patients received open-label fluoxetine hydrochloride 20 mg daily. Phase 3 was a 25-week double-blinded, continuation treatment phase. In this phase the patients identified as having responded to fluoxetine hydrochloride 20-mg daily (at phase II) were assigned by random allocation to WEKKLY-90, DAILY-20, or PLACEBO. Phase 4 was a rescue treatment phase. In this phase, the patients who were identified as having relapsed in phase 3 had their dose escalated. A total of 501 patients were randomized (at phase III) to the treatment groups WEEKLY-90 (n=190), DAILY-20 (n=189), and PLACEBO (n=122). Figure 1 lists the flowchart of patient's disposition.

The study participants were outpatients of age 18 to 80 years who met DSM-IV criteria for nonpsychotic major depression single episode or recurrent with a current episode duration of >=4 weeks of moderate intensity, confirmed by the structured clinical interview for DSM-IV, Patient version (SCID-P). Patients must have had a modified 17-item Hamilton Depression Rating Scale (HAMD17) score of >=18 and a Clinical Global Impressions of Severity (CGI-Severity) score of >=4 at the screening phase.

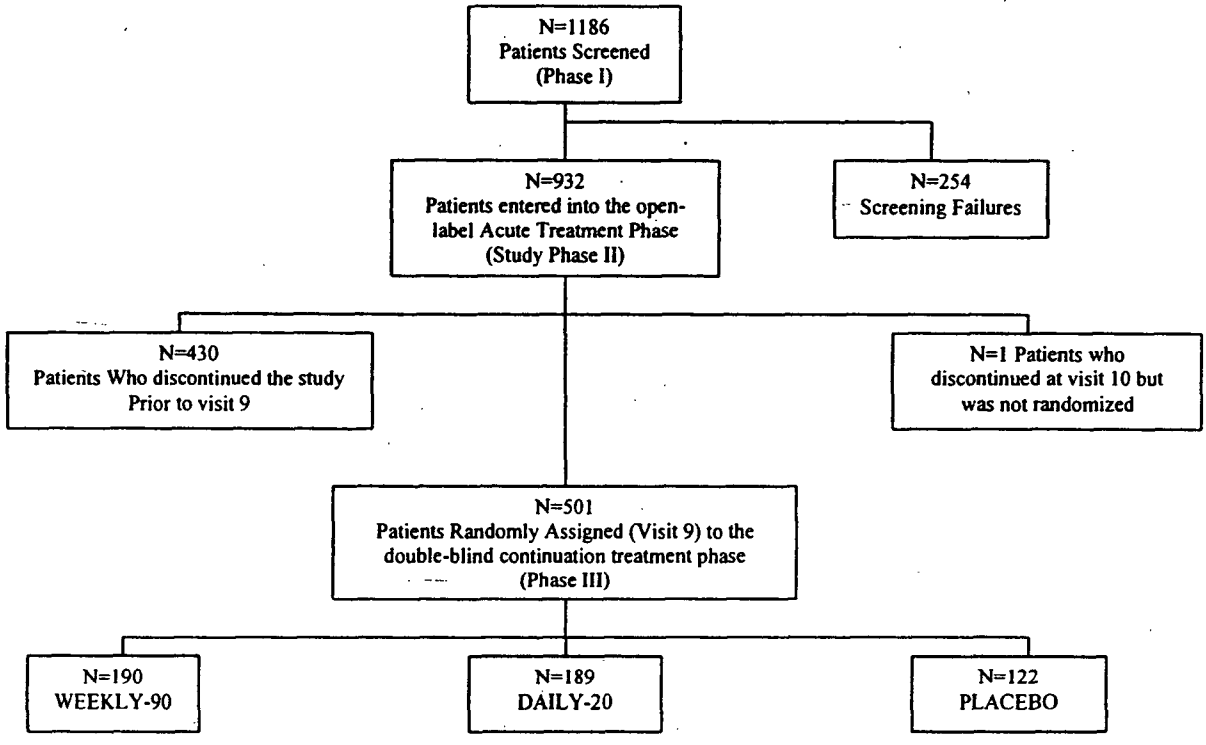

Figure 1. Flowchart of patients

There were two primary objectives in the double-blind continuation treatment period. The first primary objective was to determine the long-term antidepressant efficacy of enteric-coated fluoxetine hydrochloride 90-mg once weekly by comparing the relapse rate of patients given 90 mg fluoxetine once weekly with that of patients given placebo, after 16 weeks of continuation therapy. The second primary objective was to determine whether 90-mg fluoxetine once weekly was non-inferior to standard treatment for long-term treatment of depression. This objective was evaluated by estimating whether the relapse rate of patients given 90 mg fluoxetine once weekly was not appreciably higher (less than 15 percentage points) than those of patients continuing on 20 mg fluoxetine daily, after 16 weeks of continuation therapy.

The secondary objectives of this study were (a) to compare relapse rate of patients undergoing continuation treatment, (b) to compare the HAMD17*, other subscales from HAMD28, and CGI severity scores (change from baseline to endpoint) of patients undergoing continuation treatment, (c) to compare the rate of treatment-emergent adverse events, (d) to compare patients' perceptions of efficacy while undergoing continuation treatment. These objectives were evaluated by comparing relapse rates across the entire 25-week treatment period, change from baseline to endpoint of HAMD17*, other subscales from HAMD28, and CGI severity scores, severity of symptoms during the double-blind continuation treatment phase, rate of treatment-emergent adverse events,

and patients' perceptions of efficacy and quality of life of patients for each of the treatment conditions.

The primary efficacy criteria was the physician's categorical determination that a patient had relapsed. According to the protocol, relapse was defined as meeting the symptom criteria for major depressive episode, as determined by the Major Depressive Episode (MDE) module of the SCID-P and an increase in the CGI-severity score of >=2 over the score at visit 9 (i.e., at the time of randomization to double-blind continuation treatment phase) for two consecutive visits. If, at a patient's sixth unscheduled visit for significant reemergence of symptoms, the patient's HAMD17 score was >9 and CGI-Severity score was >2, the patient was classified as having relapsed. Secondary efficacy measures were HAMD28, HAMD17, CGI-severity, SF-36, Zung SDS, and PGI-Sexual Function.

The primary endpoint compared the proportions of patients for each treatment group who relapsed after 16 weeks of continuation treatment based on Kaplan-Meier analysis of time-to-relapse. The primary analysis was performed on an intent-to-treat population. All statements of statistical significance were based on a two-tailed test with α=0.05 unless stated otherwise. Tests for non-inferiority were based on a one-tailed test with α=0.05.

The primary analysis used confidence intervals to estimate and compare the relapse rate after 16 weeks of continuation treatment among the three treatment groups. The confidence intervals were constructed using estimates of the relapse rate and standard error for each group obtained from Kaplan-Meier analysis of the time-to-relapse. The efficacy of WEEKLY-90 compared with PLACEBO and of DAILY-20 compared with PLACEBO were assessed using 100(1-α)% two-tailed confidence intervals for the difference in relapse rates. The non-inferiority of WEEKLY-90 compared with DAILY-20 was assessed using a 100(1-α)% one-tailed confidence interval for the difference in relapse rates. The range of non-inferiority was defined as 0.15. The log-rank test was used to compare the entire time-to-relapse curves of WEEKLY-90, DAILY-20, and PLACEBO.

A secondary analysis of the time-to-relapse was conducted using a proportional hazard regression model with treatment groups, investigator, and treatment-by-investigator interaction as effects in the model. ANCOVA analyses were done in analyzing the other secondary measures (HAMD17 total score, HAMD28 subscale factors, CGI-severity, and quality of life subscales).

The primary efficacy variable, time-to-relapse, was analyzed for different subgroups (e.g., gender, age<50, age>=50, and race) by conducting proportional hazard regression analyses.

In the protocol, a planned interim analysis was proposed when approximately 50% of patients had provided data by 16 weeks of continuation treatment. The proposed interim analysis was not conducted. However, an interim analysis (an amendment occurred before the interim analysis) was conducted when all patients (100%) attained the primary endpoint or discontinued prior to this point. In fact, only 38 patients remained in the treatment continuation phase (phase III) at the time of the interim analysis. As all data relevant to the primary endpoint was included in the interim analysis, no adjustment to α(significance level) was made.

Sponsor's Results:

The results reported here are based on the 501 patients who were subsequently randomized at the double-blind continuation treatment phase (i.e. phase III). There were 68.3% females among the 501 patients. Majority (89.6%) of the patients were Caucasians. The mean age of the patients was 41.46 (range from 19-75 years) years. No statistically significant differences among the three treatment groups were observed with respect to age, gender and race. At the time of randomization, there were no statistically significant differences among the three treatment groups with respect to clinical characteristics.

The primary efficacy analysis compared relapse rates at 16-weeks post-randomization (visit 15) based on Kaplan-Meier analysis of time-to-relapse, where relapse was defined per protocol. Figure 2 lists the Kaplan-Meier survival curves of the three treatment groups. Table 1 lists the results obtained from the Kaplan-Meier survival analysis. The log-Rank test for comparing the survival curves of WEEKLY 90 vs. Placebo was statistically significant (p=.007). The test showed that the patients on WEEKLY-90 were less likely to relapse than the patients assigned to PLACEBO. However, at week 16, the difference in the relapse rates for Weekly 90 vs. placebo was not statistically significant (p=0.093), but it was significant at week 25 (p=.038). The statistical tests for testing non-inferiority of WEEKLY 90 to DAILY 20 were insignificant at weeks 16 - 25.

Table 1: Summary of Efficacy Endpoints:

	Treatment Group Comparison p-values			
Analysis	Weekly 90 vs. Placebo	Daily-20 vs. Placebo	Weekly 90 vs. Daily 20	Weekly 90 NI* to Daily 20
Relapse Rates				
Log-Rank, Weeks 1-25	.007	<.001	.164	
Kaplan-Meier, Week 16	.093	.003		.075
Kaplan-Meier, Week 25	.038	<.001		.185
Completers, Week 16	.040	.001		.122
Completers, Week 25	0.41	<.001		.299

* NI= test of non-inferiority
Source of the table: Pre NDA briefing document

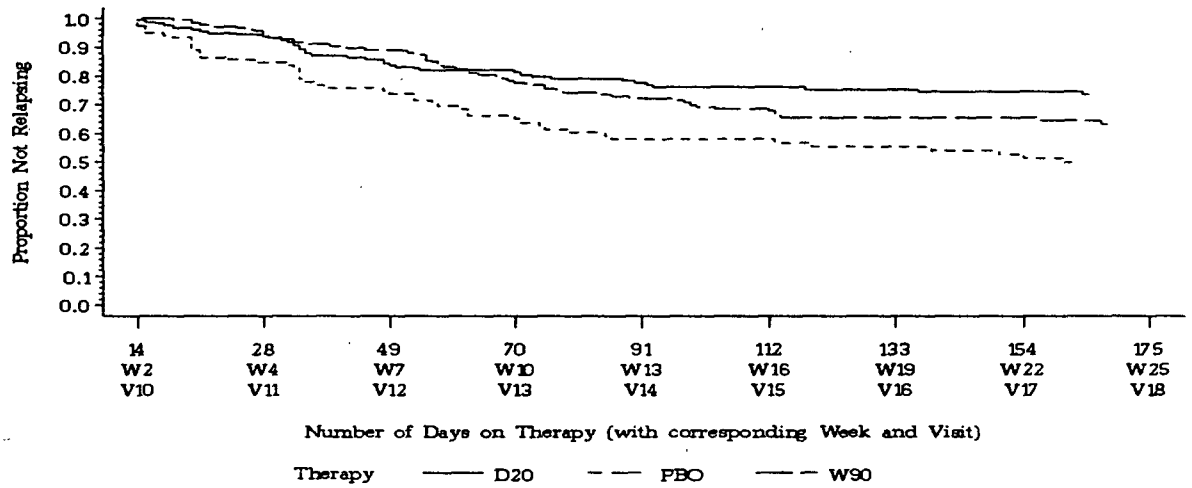

Figure 2: Kaplan—Meier Time—to—Relapse Estimates (Relapse per Protocol)

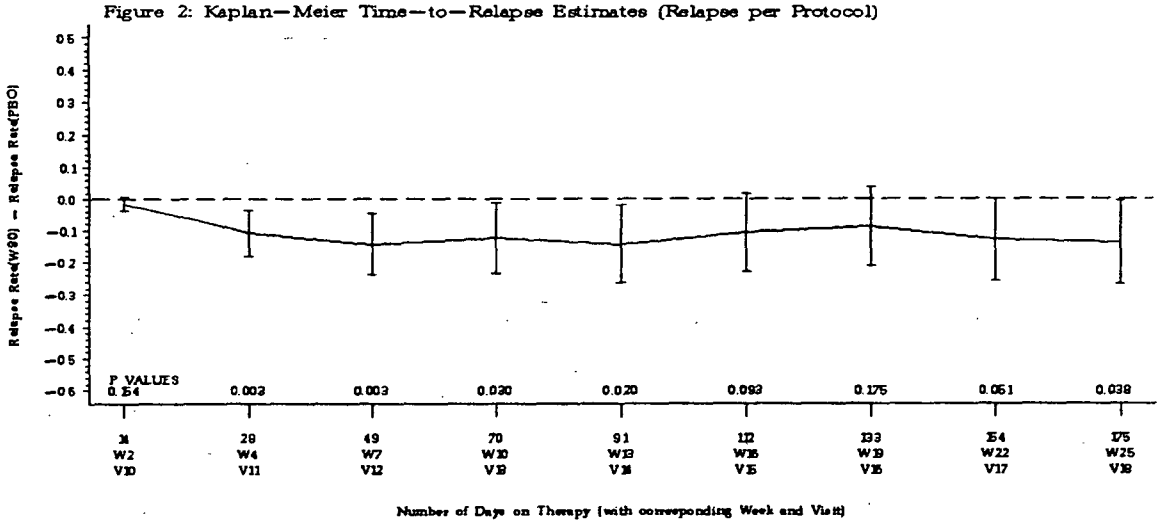

Figure 3: Two Tailed 95% CI for the Difference in Relapse Rates of WEEKLY—90 and PLACEBO

For Assessing the Superiority of WEEKLY—90 relative to PLACEBO
Entire CI should fall below 0
Based on Kaplan—Meier Estimates (Relapse per Protocol)

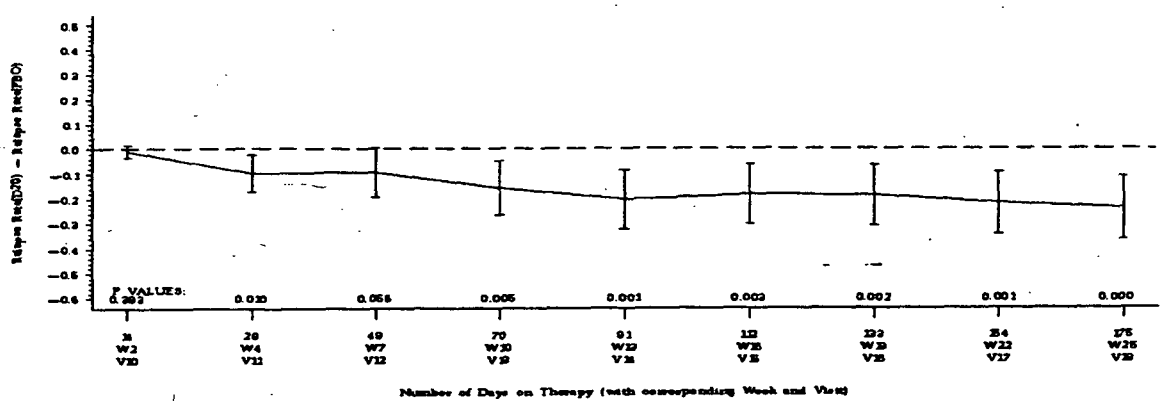

Figure 4: Two Tailed 95% CI for the Difference in Relapse Rates of DAILY—20 and PLACEBO

For Assessing the Superiority of DAILY—20 relative to PLACEBO
Entire CI should fall below 0
Based on Kaplan—Meier Estimates (Relapse per Protocol)

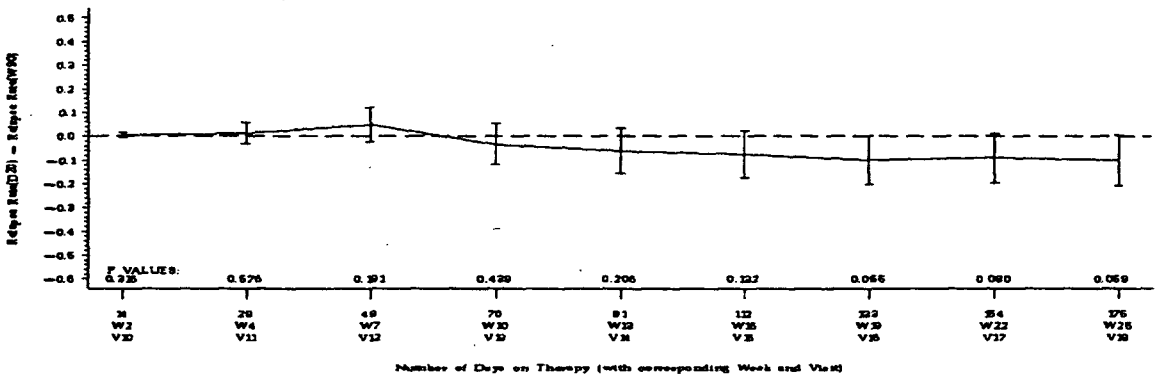

Figure 5: Two Tailed 95% CI for the Difference in Relapse Rates of DAILY−20 and WEEKLY−90

For Assessing the Superiority of DAILY−20 relative to WEEKLY−90
Entire CI should fall below 0
Based on Kaplan−Meier Estimates

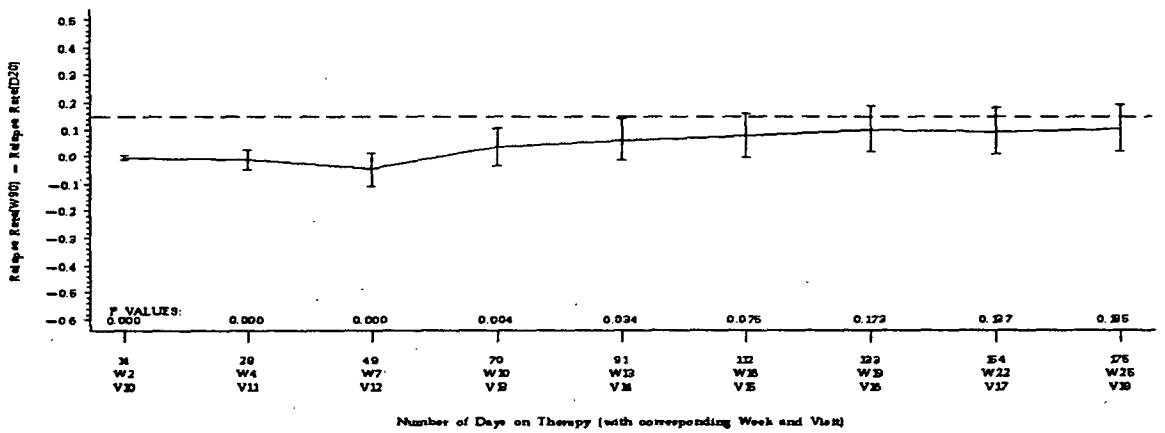

Figure 6: Upper Tailed 95% CI for the Difference in Relapse Rates of WEEKLY−90 and DAILY−20

For Assessing the Non−inferiority of WEEKLY−90 relative to DAILY−20
Upper Limit of CI should fall below 0.15 (range of non−inferiority)
Based on Kaplan−Meier Estimates (Relapse per Protocol)

Figures 3 lists the 95% confidence interval plots to assess the superiority and non-inferiority of WEEKLY-90 relative to DAILY-20 and PLACEBO. The estimated relapse rates for patients on WEEKLY-90 were always less than that for patients on PLACEBO. Up to Week 13, the difference between the relapse rates for WEEKLY-90 and PLACEBO were statistically significant (P<.05). But from week 16 and onwards, the 95% confidence intervals tended to include zero in the intervals. Figure 4 indicates that the relapse rates for DAILY-20 group were significantly (p<.05) less than the rates for PLACEBO group and consistent from week 7 to week 25. Figure 5 indicates that WEEKLY-90 and DAILY-20 were not different up to week 13, but from week 16 and onwards the two groups tended to be different (although the two groups were not statistically significant at p=.05).

Figure 6 assesses the non-inferiority of WEEKLY-90 relative to DAILY-20. The non-inferiority of WEEKLY-90 relative to DAILY-20 was confirmed if the entire confidence interval (95%, one-tailed) fell below .15 (range of non-inferiority was defined

in protocol). Non-inferiority was demonstrated for 3 months. After 3 months, the upper end of the confidence interval fell above .15 and non-inferiority was not demonstrated.

The secondary analyses results based on proportional Hazards model also agreed closely with the results based on the log-rank test. The conclusions drawn based on the alternative definition of relapse (i.e., Physician's determination of relapse) were also consistent with the analyses based on the definition of relapse as specified in the protocol. Approximately 93% of patients classified by physicians as relapse agreed with the protocol's definition of relapse. Relapse rates for patients assigned to WEEKLY-90 or DAILY-20 were lower than those compared with patients on PLACEBO, and the relapse rates for patients on WEEKLY-90 and DAILY-20 did not differ significantly. A completers [1] analysis results agreed closely with those based on the Kaplan-Meier analysis.

The analyses of change from baseline to endpoint (last observation carried forward) for several efficacy measures [2] indicated that significantly greater worsening occurred for patients assigned to PLACEBO relative to WEEKLY-90 and DAILY-20. The two active treatment groups were not statistically significantly different in the analyses.

Subgroups were not analyzed separately but were examined by including corresponding covariates (age: <=50, >50, Gender, Race) and covariate-by-treatment interactions in the proportional hazards regression model of time-to-relapse. There were no statistically significant demographic covariate-by-treatment interaction effects.

Adverse Events:

Of the 501 patients (randomized at period III), 370 (73.9%) experienced at least one treatment-emergent adverse event. The most frequently occurring treatment-emergent adverse events were headache (10.8%), nervousness (10.4%), rhinitis (9.8%), somnolence (9.2%), and asthenia (9.0%). Fourteen (2.8%) patients discontinued due to an adverse event. The most commonly reported treatment-emergent adverse events among patients treated with WEEKLY-90, DAILY-20, and PLACEBO were comparable. There was no death in this study.

Sponsor's Final Conclusion:

In the Kaplan-Meier analysis of time-to-relapse as the primary method of comparison, the relapse rate of patients on PLACEBO was always higher than that for patients on either WEEKLY-90 and DAILY-20. The Kaplan-Meier time-to-relapse curves for WEEKLY-90 and DAILY-20 were highly similar for the initial 2 to 3 months; thereafter the relapse rate for WEEKLY-90 was intermediate to the relapse rate for DAILY-20 and PLACEBO. At the end of continuation treatment patients on WEEKLY-90 had significantly a lower relapse rate compared with patients on PLACEBO (p=.038) although at the pre-defined

[1] Includes only those patients who either relapsed prior to, or completed to a specified endpoint.
[2] HAMD core, HAMD Subscale 5, HAMD-Anxiety Total, HAMD item1, and HAMD17*, CGI-Severity.

endpoint (16 weeks post-randomization) the superiority of WEEKLY-90 over PLACEBO was not maintained (p=.093). The non-inferiority of WEEKLY-90 relative to DAILY-20 was not established using the .15 as the declared threshold for non-inferiority (p=.075) at week 16).

Reviewer's Analysis and comments:

This reviewer reanalyzed the data set according to the statistical plan specified in the protocol. The findings were consistent with the sponsor's reported findings. This was true for both primary and secondary outcome measures. The reviewer was also able to reproduce the figures 2-6, and these are exactly same as the figures provided by the sponsor.

Reviewer's Overall Conclusion:

In this new drug application, the sponsor designed the trial and analyzed the dataset appropriately to assess the long-term efficacy of the 90-mg weekly dose of fluoxetine as compared to placebo and 20-mg daily dose of fluoxetine for long-term treatment of depression. The survival curve for WEEKLY-90 was different from the curve for PLACEBO (p=.007, log-rank test). The visit-wise comparison of relapse rates based on the estimates from Kaplan-Meier analysis demonstrated that WEEKLY-90 was efficacious as compared to PLACEBO and it was non-inferior to DAILY-20 for up to 13 weeks. After week 13, the relapse rate for WEEKLY-90 group was intermediate to the relapse rates for DAILY-20 and PLACEBO. From week 16, the visit-wise confidence intervals (in figure 3) demonstrated that the rate for WEEKLY-90 was not statistically significantly (except at week 25) different from the rate for PLACEBO. Figure 5 demonstrated that the relapse rates for WEEKLY-90 were higher (although not statistically significant at p=.05) as compared to the corresponding rates for DAILY-20. Figure 6 demonstrated that WEEKLY-90 was non-inferior to DAILY-20 for first 3 months, after 3 months, WEEKLY-90 failed to demonstrate the non-inferiority to DAILY-20.

From the above findings it could be concluded that WEEKLY-90 was efficacious as compared to PLACEBO and had the same levels of efficacy as compared to DAILY-20 up to 13 weeks for treatment of depression. But at the later weeks, WEEKLY-90 was less effective as compared to DAILY-20. This reviewer found sufficient evidence from the statistical analyses of this clinical trial data set to support the sponsor's first claim that WEEKLY-90 was efficacious as compared to PLACEBO for treatment of depression. This reviewer did not find sufficient evidence to support the second claim that WEEKLY-90 was non-inferior to DAILY-20 for long-term treatment of depression.

The statistical methodology used to evaluate the non-inferiority of a drug in this application seems to be questionable. Therefore, it is suggested not to refer this application in the future non-inferiority trials.

[/S/]

Ohidul Siddiqui, Ph.D
Mathematical Statistician

Concur: Dr. Kun Jin

Dr. George Chi

CC:
Arch NDA # 21-235
HFD-120/Dr. Katz
HFD-120/Dr. Laughren
HFD-120/Dr. Kathy Smith
HFD-120/Mr. David
HFD-344/Dr. Barton
HFD-710/Dr. Chi
HFD-710/Dr. Jin
HFD-710/Dr. Siddiqui
HFD-710/Chron