RECEIVED MAR 1 3 1998

Date: MAR 9 1999

Statistical Review and Evaluation

NDA #: 20-822

Applicant: Forest Laboratories, Inc.

Name of the Drug: Citalopram Hydrobromide Tablets

Indication: Treatment of Depression

Documents Reviewed: Volumes 1.323 to 1.488, amendments dated 7-31-97, 8-14-97, 9-4-97, 10-09-97,10-24-97, 11-7-97, 11-18-97, 11-21-97, 12-12-97, 01-02-98, 02-23-98

Clinical Reviewer: Gregory Dubitsky, M.D. (HFD-120)

The issues in this review have been discussed with the reviewing medical officer, Dr. Gregory Dubitsky, M.D. (HFD-120).

Various Sections of this review are:

I. **Background/Introduction**

According to the sponsor, the so-called "Group 1" studies form the core of the efficacy and safety data from the clinical development program that will be used to support the claims being made for use of citalopram in depressed patients.

The 19 clinical trials classified as Group 1 studies, are organized into 4 main subgroupings. Three of these subgroupings are comprised of 17 studies in moderate to severe depression and are discussed in the ISE: [1] 9 placebo-controlled, short-term, domestic or foreign studies; 2 longer term, foreign studies; [2] 6 active-controlled, 5 short-term and one longer term, foreign studies using a serotonin selective reuptake inhibitor [SSRI] or a tricyclic antidepressant [TCA] as an active antidepressant comparator drug; [3] 2 uncontrolled studies (longer term, one domestic and one foreign)

A one-page overview is in the attached Table 0.1.1. This reviewer has individually reviewed the two short-term studies which the sponsor claimed to provide statistically significant evidence of efficacy and the two long-term studies. An overview of results of all [including, 4 (5 for CGI severity of illness) short-term studies, which the sponsor was interested to provide] studies are in the first ("0") Section of the appendix.

All the 147 (105 on citalopram and 42 on placebo) patients in one long-term Study 89305 were derived from citalopram patients in studies 89303 and 89306. Counting these patients twice, there were 1320 patients on citalopram and 562 patients on placebo in the double-blind phase of the 9 placebo-controlled studies (first subgroup). The numbers of patients in the citalopram group of the two studies 86A and 87A were only 20 (placebo 4) and 16; respectively. The sponsor did not provide results for these two studies in the comprehensive tables and confidence intervals (Nov. 7, 1997 submission).

Foreign Marketing History

According to the sponsor, as of December 31, 1996, citalopram is marketed for depression in 49 countries, and marketing applications are pending in 23 countries. The dosage formulations marketed include tablets and intravenous solution.

No foreign regulatory authority has refused approval of citalopram on safety grounds. There have been no withdrawals of citalopram from marketing by foreign authority for any reason related to safety or efficacy. The applicant, Lundbeck, withdrew citalopram 10mg in France before obtaining marketing authorization due to insufficient efficacy. There have been no reports obtained from foreign regulatory authorities of adverse effects or major changes in marketing status.

II. **Clinical Studies**

All analyses referred to in this report are the sponsor's analyses, except where specifically mentioned to be done by this reviewer.

This reviewer consulted Dr. Dubitsky (HFD-120) regarding the most important efficacy variables. They are "Change from Baseline in HAM-D Total", "Change from Baseline in HAM-D Depressed Mood Item", "Change from Baseline in CGI Severity of Illness", and "Change from Baseline in MADRS Total," for the short-term studies.

1. **Study 85A**

Study 85A was a randomized, 4-week double-blind, placebo-controlled, 20 to 80 mg/day citalopram flexible-dose, three-center U.S. study consisting of a 1-week, single-blind, placebo washout period, in outpatients (180 enrolled and 169 ITT patients) with major depression or bipolar disorder, depressed.

1A. Objective

To assess the comparative safety and efficacy of citalopram and placebo in moderately to severely ill patients with major depression or bipolar disorder, depressed.

1B. Disposition of Patients

In the citalopram group, 48% patients withdrew, of which 25% withdrew because of adverse events and 7% because of lack of efficacy. In the placebo group, 44% patients withdrew, of which 8% withdrew because of adverse events and 25% because of lack of efficacy.

More citalopram (n=13) than placebo (n=4) patients discontinued therapy within the first week. During later weeks, incidences of discontinuation were comparable between the two groups (slightly more discontinuation from the placebo group).

For the intent-to-treat patients, the in-study patient numbers differed by 8% at Weeks 2 and 3 but equaled (59%) at Week 4 (submission of August 14, 1997).

Percents of patients (1) not dropping out due to lack of efficacy and (2) continuing in the study (after dropping out due to all

reasons) are graphed in the attached Figure 1.2.1.

1C. Comparability of Treatment Groups

The sponsor stated (this reviewer does a quick verification through the Tables provided by the sponsor), "The groups were similar with respect to race, age, body weight, height, and baseline vital signs; marital status, educational level, and general occupation category; and patient diagnosis."

In the Intent-to-Treat patients set, the percentage of females varied as: 32% (placebo) and 35% (citalopram).

The sponsor noted between-treatment differences in two categories of medical/surgical history: Gastrointestinal history and genitourinary history. The rate of positive history in these categories was twice as high in the placebo group (15% and 30%, respectively) as in the citalopram group (7% and 18%, respectively).

All the p-values provided for baseline comparison with respect to HAMD 24-item score, CGI severity score, HAMD 17-item score, or Zung total score, were large (non-significant).

1D. Efficacy Results (Sponsor's Analyses)

The protocol was signed by investigators on 2/13/84. This protocol did not mention the statistical methods. The sponsor stated, "Between-group effects were tested with analysis of covariance (ANCOVA) models with baseline value as covariate and including the treatment (TG) and Center (C) main effects and the TG by C interaction effect."

On request, the sponsor justified the inclusion of the interaction term in the final analysis model by saying, "We thought the minor disparity of sample size among the three centers would have negligible effect on the statistical results at most," and provided results also excluding the interaction term. There were, indeed, no major differences. The p-values were smaller in many cases, by exclusion of the interaction term.

The (1) Results with p-values and (2) Graphs for cumulative distribution functions, for (adjusted) Mean Changes From Baseline are attached as Table 1.3.1 and Figure 1.3.2 (HAM-D Total), 1.4.1 (HAM-D Depressed Mood Item), 1.5.1 and 1.5.2 (CGI Severity of Illness).

Following are the citalopram vs placebo mean differences (-ve sign indicates superiority of citalopram) and p-values.

STUDY 85A

HAMILTON DEPRESSION SCALE - 24-ITEM TOTAL

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

WEEK	Citalopram Vs Placebo Difference	P-value
1	-2.47	0.0116*
2	-3.77	0.0027*
3	-2.86	0.0565
4	-3.32	0.0344 *

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

WEEK	Citalopram Vs Placebo Difference	P-value
1	-2.62	0.0086*
2	-5.72	0.0001*
3	-4.22	0.0159*
4	-4.18	0.0250 *

* Statistically Significant at 5% level

HAMILTON DEPRESSION SCALE - Depressed Mood ITEM

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

WEEK	Citalopram Vs Placebo Difference	P-value
1	-0.33	0.0044*
2	-0.42	0.0021*
3	-0.34	0.0234*
4	-0.49	0.0023*

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

WEEK	Citalopram Vs Placebo Difference	P-value
1	-0.35	0.0026*
2	-0.50	0.0006*
3	-0.39	0.0240*
4	-0.55	0.0024 *

* Statistically Significant at 5% level

CGI-Severity Score

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

WEEK	Citalopram Vs Placebo Difference	P-value
1	-0.15	0.0824
2	-0.34	0.0024*
3	-0.24	0.0650
4	-0.26	0.0508 (Stated to be .046* In 12-12-97 submission)

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

WEEK	Citalopram Vs Placebo Difference	P-value
1	-0.16	0.0734
2	-0.52	0.0001*
3	-0.28	0.0645
4	-0.34	0.0373 *

* Statistically Significant at 5% level

These results provided clear statistical evidence in favor of the efficacy of citalopram, except that CGI Severity results were slightly weaker (in particular, at Week 3) and HAM-D Total LOCF result at Week 3 was marginally non-significant.

Dropout Cohorts

Mean HAM-D Total scores and Mean Changes From Baseline for subgroups of patients dropping out at different times and for completers are in Figures 1.3.3 and 1.3.4. Mean Changes From Baseline for subgroups of patients by reason of Withdrawal and for completers are in Figure 1.3.5. In Figure 1.3.3, irrespective of time of dropout or completion, placebo scores were inferior. In Figure 1.3.4, among the patients who were included in Week 1 dropouts, the placebo patients had better changes from baseline than those of the citalopram patients (contradiction with the previous statement occurs because of baseline differences in this group of dropouts). This is likely to inflate (we cannot say that wrongly inflated) the later OC results in favor of citalopram and LOCF results in favor of placebo. However, the opposite (and larger) for those who were included in Week 2 dropouts will more than offset this, after Week 2. Therefore, there is no special concern about interaction of treatment with this factor (completion or dropping out), in spite of the following observation in another type of classification of dropouts.

Of the patients dropping out due to Adverse events and LOE & AE, the placebo patients had better changes from baseline (HAM-D Total) than those of the citalopram patients (actually, deterioration for citalopram patients dropping due to LOE & AE).

Mean CGI Severity scores and Mean Changes From Baseline for subgroups of patients dropping out at different times are in Figures 1.5.3 and 1.5.4. Mean Changes From Baseline for subgroups of patients by reason of Withdrawal and for completers are in Figure 1.5.5. The findings are practically the same as for HAM-D total.

1E. Reviewer's Comments and Conclusions on Study 85A

Based on the sponsor's submitted results, Study 85A provided statistical evidence in favor of the efficacy of citalopram, although, in a few cases, the results were marginally non-significant. This reviewer's analyses by 2-sample Wilcoxon test provided similar evidence.

The results across the three sites were reasonably consistent.

The average daily dose (page 10-00414 of NDA) of citalopram during Week 1 was 31 mg/day. It increased to 54mg/day during Week 2, and to around 60 mg/day during Weeks 3 and 4. For comparative purposes, the dose of placebo was calculated by assuming 20mg per tablet. The Mean Daily Dose for the placebo group was greater than that for the citalopram group starting from Week 2 and the difference reached statistical significance during Weeks 3 and 4.

On a request for clarifications, the sponsor stated, "In the NDA table, the total dose ingested for a specific week was divided by the number of days between two consecutive visits to obtain the mean daily dose for that week. While in the table provided on 08/14/97, the total dose ingested was divided by the number of days the patients actually took study medication." The previous paragraph relates to the NDA table.

There was a significant age-by-treatment interaction in this study because for patients greater than 60 years old, placebo (n=4) did better than citalopram (n=7). This may alert us for all such findings. However, with such a small number of patients in this age group, we cannot confirm that citalopram is inefficacious or worse than placebo in this age group.

2. **Study 91206**

Study 91206 was a randomized, 6-week double-blind, placebo-controlled, fixed-dose (10,20,40, or 60 mg/day citalopram), twelve-center U.S. study consisting of a 1-week, single-blind, placebo washout period, in outpatients (650 enrolled and 605 ITT patients) with moderate to severe depression.

2A. Objective

To demonstrate safety and efficacy of citalopram by comparing four doses (10, 20, 40, or 60 mg/day) to placebo in male and females outpatients meeting DSM-III-R criteria for major

depressive episode and, additionally, to estimate the minimum effective dose.

2B. Disposition of Patients

Thirty-three percent of placebo and 33% of citalopram patients failed to complete the study. A total of 15% (79/521) citalopram-treated patients discontinued because of adverse events compared to 6% (8/129) placebo-treated patients ($p<0.004$ for overall difference between treatment groups). There was a trend for a between-group difference for discontinuations due to unsatisfactory therapeutic response ($p=0.053$). Four percent (19/521) citalopram patients discontinued for lack of efficacy compared to 9% (11/129) placebo-treated patients.

For the intent-to-treat patients, the in-study patient numbers differed most among treatment groups at Week 2 (by 8% among the doses, placebo was not at the extreme; Amendment of August 14, 1997).

Percents of patients (1) not dropping out due to lack of efficacy and (2) continuing in the study (after dropping out due to all reasons) are graphed in the attached Figure 2.2.1.

2C. Comparability of Treatment Groups

The sponsor stated (this reviewer does a quick verification through the Tables provided by the sponsor), "The groups were similar with respect to age, body weight, height ... marital status, histories of drug and alcohol abuse."

In the Intent-to-Treat patients set, the percentages of females in the 5 treatment groups varied between 53% (citalopram 60mg) and 66% (citalopram 20mg). The pairwise comparison between these two groups provided a p-value of .043 (done by the reviewer). Since placebo is not involved, this imbalance may not, generally, be of any concern. In spite of a greater proportion of female patients in the 20 mg group and female patients doing slightly better (page 8- 73905) than male patients in this 20mg (not in 40 or 60 mg), 20mg results in this study was poor (indicating a really poor performance for 20 mg in this study). The sponsor provided the p-value only for the overall comparison ($p=.183$).

Alcohol abuse varied from 5% (10mg Citalo.) to 14% (20mg Citalo.).

The between-group difference with respect to the ongoing use of tobacco was statistically significant ($p=.002$).

All the p-values provided for baseline comparison with respect to HAMD 21-item total, MADRS, SCL-56, or factor scores within each, were large (non-significant).

2D. Efficacy Results (Sponsor's Analyses)

The protocol stated, "Analysis of covariance, with treatment and center as factors and baseline score as covariate, will be performed."

The (1) Results with p-values and (2) Graphs for cumulative distribution functions, for (adjusted) Mean Changes From Baseline, are attached as Table 2.3.1 and Figure 2.3.2 (HAM-D Total), 2.4.1 (HAM-D Depressed Mood Item), 2.5.1 and 2.5.2 (CGI Severity of Illness), and 2.6.1 and 2.6.2 (MADRS).

Following are the citalopram doses vs placebo mean differences (-ve sign indicates superiority of citalopram) and p-values.

STUDY 91206

HAMILTON DEPRESSION SCALE - 21-ITEM TOTAL

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	-0.21	.718	+0.09	.890	-0.07	.896	-0.02	.887
2	-0.80	.261	+0.25	.719	-0.45	.528	+0.22	.717
3	-0.30	.712	+0.28	.733	-0.98	.231	-0.40	.635
4	-0.99	.246	-0.20	.816	-1.54	.074	-1.69	.055
5	-1.22	.188	-0.40	.663	-1.86	.022	-1.95	.042
6	-1.47	.124	-0.63	.506	-2.92	.003*	-2.74	.005*

* Clearly significant at .05 level, by multiple comparison adjustments

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	-0.08	.898	+0.20	.732	+0.01	.982	+0.12	.837
2	-0.90	.215	+0.32	.658	-0.77	.298	0.00	.998
3	-0.90	.284	+0.13	.885	-1.26	.133	-0.60	.488
4	-1.59	.075	0.00	.994	-1.86	.037	-1.40	.126
5	-1.87	.056	+0.23	.817	-1.80	.066	-1.16	.255
6	-1.25	.213	+0.61	.546	-1.99	.050	-1.61	.126

The last observation carried forward (LOCF) results showed the efficacy of citalopram 40 and 60 mg at Week 6. The corresponding results for observed cases (OC) were not only non-significant, the numerical benefits over placebo were also smaller. Although these results were not statistically strong, the numerical benefit of citalopram was shown for the higher doses, starting from Week 3.

HAM-D Depressed Mood Item (1)

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	-0.16	.106	-0.16	.094	-0.24	.015	-0.21	.043
2	-0.27	.024	-0.25	.035	-0.24	.042	-0.32	.009*
3	-0.23	.062	-0.23	.059	-0.33	.008*	-0.36	.005*
4	-0.25	.066	-0.21	.103	-0.34	.012*	-0.48	.001*
5	-0.35	.014*	-0.28	.048	-0.53	.001*	-0.52	.001*
6	-0.43	.003*	-0.35	.016	-0.58	.001*	-0.68	.001*

* Clearly significant at .05 level, by multiple comparison adjustments

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	-0.16	.117	-0.16	.117	-0.26	.150?	-0.20	.056
2	-0.28	.018	-0.19	.102	-0.26	.030	-0.29	.017
3	-0.22	.078	-0.19	.147	-0.39	.002*	-0.37	.005*
4	-0.28	.047	-0.09	.513	-0.37	.011*	-0.45	.003*
5	-0.41	.006*	-0.18	.224	-0.58	.001*	-0.41	.009*
6	-0.46	.003*	-0.20	.202	-0.49	.002*	-0.58	.001*

* Clearly significant at .05 level, by multiple comparison adjustments

With respect to HAM-D depressed mood item, staistical evidence for the efficacy of citalopram is strong, although such evidence is absent for 20mg (especially, in the OC analysis). Numerical benefits for 20mg over placebo were always larger in the LOCF analysis than those in the OC analysis, right after the first week. [HAM-D depressed mood item scores for dropout cohorts have not been provided.]

CGI-Severity Score

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	+0.03	.694	+0.06	.694	+0.04	.557	+0.01	.905
2	-0.07	.342	0.00	.980	-0.03	.674	+0.05	.499
3	+0.03	.758	-0.03	.699	-0.05	.561	-0.02	.748
4	-0.02	.817	+0.02	.862	-0.16	.064	-0.10	.253
5	-0.02	.839	-0.03	.760	-0.19	.056	-0.10	.347
6	-0.06	.609	-0.03	.743	-0.21	.049	-0.16	.132

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	+0.03	.634	+0.07	.323	+0.03	.661	+0.02	.828
2	-0.10	.194	+0.01	.935	-0.04	.618	+0.07	.408
3	+0.01	.942	-0.04	.640	-0.06	.487	-0.04	.656
4	-0.07	.456	+0.05	.630	-0.17	.067	-0.07	.478
5	-0.08	.424	+0.02	.856	-0.17	.112	-0.01	.971
6	-0.10	.376	+0.16	.167	-0.12	.290	-0.06	.609

With respect to CGI Severity, there is no statistical evidence in favor of the efficacy of citalopram. Even the numerical superiority of citalopram over placebo is negligible.

MADRS TOTAL Score

LOCF

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	-0.47	.529	+0.02	.978	-0.48	.519	-0.50	.515
2	-0.88	.335	+0.66	.463	-0.62	.502	-0.82	.383
3	-0.62	.551	-0.18	.860	-1.72	.101	-1.82	.089
4	-1.46	.192	-0.64	.564	-2.25	.045	-2569	.026
5	-1.81	.130	-1.15	.330	-3.70	.002*	-3.20	.010*
6	-2.43	.053	-1.66	.180	-4.17	.001*	-4.38	.001*

* Clearly significant at .05 level, by multiple comparison adjustments

OBSERVED

(With Respect to MEAN CHANGE FROM BASELINE)

	Cit.10 Vs Pla		Cit.20 Vs Pla.		Cit.40 Vs Pla.		Cit.60 Vs Pla.	
WEEK	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value	Diff.	P-Value
1	-0.44	.555	+0.07	.928	-0.63	.409	-0.45	.560
2	-1.05	.249	+1.05	.255	-0.74	.432	-0.61	.518
3	-0.95	.373	+0.10	.927	-1.95	.070	-2.09	.062
4	-2.37	.042	-0.21	.856	-2.47	.035	-2.54	.038
5	-2.75	.025	-0.58	.643	-3.65	.003*	-2.27	.077
6	-2.36	.048	-0.14	.920	-3.03	.025	-3.01	.032

* Clearly significant at .05 level, by multiple comparison adjustments

The last observation carried forward (LOCF) results showed the efficacy of citalopram 40 and 60 mg at Weeks 5 and 6, with respect to MADRS. The corresponding results for observed cases (OC) were not only non-significant (except for 40mg at Week 5) by multiple comparison adjustments, the numerical benefits over placebo were also smaller. Citalopram patients dropping out after Week 4 and Week 5 had a few times more improvements compared with the corresponding (dropping at those weeks) placebo patients (NDA pages 10-07654 and 10-07655). Carrying these observations forward inflated (compared with OC) the results in favor of citalopram. We do not know the exact truth about the situation to confirm if this inflation was inappropriate or not.

However, the evidence for efficacy of citalopram 40 and 60 mg (10mg, marginally) with respect to MADRS is reasonably acceptable.

Dropout Cohorts

Mean HAM-D Total scores and Mean Changes From Baseline for subgroups of patients dropping out at different times and for completers are in Figures 2.3.3 to 2.3.6 (unadjusted). Mean Changes From Baseline for subgroups of patients by reason of Withdrawal and for completers are in Figures 2.3.7 and 2.3.8.

40mg

Although those who were included as dropouts from placebo at Weeks 1 and 3 had good improvements, the improvement for those who were included as dropouts from citalopram at Week 5 was so large that, carrying these last observations forward inflated (we cannot claim that wrongly so) the efficacy of citalopram at Week 6 by LOCF method (compared with OC method).

On the other hand, the OC results were inflated in favor of citalopram up to Week 5 (the above fact outweighed this after Week 5) because those who were included as dropouts from citalopram at Week 2 had outstandingly bad HAM-D Total score at baseline (Figure 2.3.3).

Among those dropouts due to lack of effect at Week 1, 40 mg citalopram patients really deteriorated a lot whereas the placebo

patients had a small improvement (Figure 2.3.7). In other categories (of reason for dropout) citalopram patients had better improvement compared with placebo patients.

Overall, we cannot say that placebo performed better among the dropouts.

60mg

Although dropouts from placebo at Week 1 had better improvements than those from citalopram, the improvements for dropouts at Weeks 3 and 4 were so large that, carrying these last observations forward inflated (cannot say if totally wrongly) the efficacy of citalopram at Week 5 by the LOCF method (compared with the OC method). The OC results were inflated in favor of citalopram up to Week 3 because of the dropouts at Week 1. The very good placebo dropout response at Week 5 should not have any effect other than balancing some of the bigger effects of dropouts at Weeks 3 and 4.

Placebo patients dropping out due to Non Compliance had better changes from baseline (HAM-D Total) than those of the citalopram patients (Figure 2.3.8.). In other categories (of reason for dropout) citalopram patients had better improvement compared with placebo patients.

Overall, we cannot say that placebo performed better among the dropouts.

2E. Reviewer's Comments and Conclusions on Study 91206

Study 91206 provided statistical evidence in favor of the efficacy of citalopram (generally, 40 and 60 mg, and 10mg only with respect to HAM-D Item 1), strongly with respect to HAM-D Depressed Mood Item, moderately with respect to MADRS, marginally with respect to HAM-D 21 Item Total, and none with respect to CGI Severity of Illness.

The protocol elaborated on the statistical methods and stated, "No adjustment of significance level will be carried out in order to account for the multiplicity of comparisons with placebo. However, William's one-sided test of the null hypothesis ..." By the protocol mentioned primary efficacy analysis, this study provided clear statistical evidence in favor of the efficacy of 40 and 60 mg citalopram. However, the clinical and statistical reviewers tried to follow the procedures they usually follow for anti-depressants; multiple comparison adjustments and robustness

of results across LOCF and OC have played roles in the above conclusion.

Endpoint results provided on pages 8-05819 and latter in the NDA by alternative analyses (one or two of the following methods for a particular efficacy variable: William's test, ANCOVA, nonparametric, logistic, Wilcoxon Rank Sum test, Kruskall-Wallis test, Fisher's exact test for responses) provide much better impression about the efficacy of 40 and 60 mg citalopram.

Even by excluding an investigator on the disbarred list (analyses provided in July 31, 1997 submission), there does not seem to be any concern with respect to the above conclusion.

Out of the twelve centers (submission of August 14, 1997), a few centers had placebo results better than the citalopram (higher doses) results. For 20 mg, less than 50% centers had citalopram (20mg) results better than the placebo results (overall results also were the poorest for this dose). It does not seem that the overall efficacy of the higher doses was driven by a few big centers.

3. **Study 89304** (Long-term)
(Original Report Vol. 1.374, Short Revised Report in Vol. 1.364)

This was a randomized, double-blind, multicenter, placebo-controlled (Phase III), parallel group study to determine whether patients who responded to short term (8 weeks) citalopram (10 to 60 mg/day) treatment would demonstrate a therapeutic benefit in maintenance therapy (an additional 24 weeks), compared to short-term citalopram responders treated with placebo for the next 24 weeks.

Baseline values were recorded at the end of the open-label evaluation period. Visits were scheduled at the end of each week during the double-blind (also during the open-label) evaluation period.

The primary efficacy measurement was time to relapse. Assessments were made at Weeks 4, 8, 12, 16, 20, and 24. Relapse occurred if the investigator withdrew the patient from the study due to deterioration and the MADRS total score was ≥25.

Three hundred ninety-one patients were enrolled in study Period A for open-label citalopram treatment. All 391 patients took medication, with a total of 249 (66%) completing Period A. Of

the 391 Period A patients, 226 (58%) continued in Period B.

Of the 226 eligible patients from Period A constituting the intent-to-treat (ITT) population for the double-blind period, 152 (67%) were randomized to continue citalopram treatment, and 74 (33%) were randomized to receive placebo. Ninety-two (61%) citalopram and 38 (51%) placebo patients completed the trial.

A total of 31 centers (52 investigators) entered patients. In Period B, the number of patients across centers varied from 1 (center 37) to 18 (center 29).

The percent of females, in the double-blind period, were 51 (69%) in the placebo group and 111 (73%) in the citalopram group (no statistically significant imbalance between treatment groups).

The mean age of patients was 46.9 years in the placebo group and 48.6 years in the citalopram group.

Race was not recorded for the two long-term European studies. The sponsor stated, "(Virtually all patients were white)."

There were no statistically significant differences across groups when MADRS scores at entry of period B were tested (p=.31).

Efficacy Results (Sponsor's Analyses)

The study protocol originally stated that the primary measure of interest would be the difference in relapse rates. However, the analysis plan was subsequently amended; the primary measure of efficacy was to be Time to Relapse.

For the ITT population, there were 21 (13.8%) relapses in the citalopram group and 18 (24.3%) relapses in the placebo group. The sponsor stated, "There was a statistically significant lower incidence of relapse in the Citalopram-treated patients compared to placebo-treated patients (Kaplan Meier; p=0.04). ... Although the number of patients receiving each possible dose of citalopram was too small to submit to statistical analysis, there did not appear to be a dose effect in terms of relapse."

The chi-square test performed by the reviewer on the crude Relapse Rates provided a p-value of .05.

Cumulative percentages of relapse over time (4-weekly assessments) for three dose groups - citalopram up to 60mg,

citalopram 20mg, and citalopram 40mg - and placebo were provided in the Nov. 18, 1997 submission. Indeed, there did not appear to be a dose effect. However, mean number of tablets per day or week for the two treatment groups was not provided by the sponsor.

The survival curve of Time to Relapse is attached as Figure 3.3.1.

Although the primary efficacy variable provided statistically significant evidence for the maintenance efficacy of citalopram, there were only numerical and no statistically significant differences between the two treatment groups with respect to HAM-D Total, MADRS Total, or CGI Severity of Illness. This conclusion is based on the sponsor's reports discussing Week-24 ITT LOCF results. In another place (for example, for Study 89305), the sponsor stated, "... as time goes by all patients with low MADRS scores remain in the study. Because of this, and in order to avoid bias, the secondary analysis of MADRS scores was based upon ITT LOCF."

Since relapses (considering MADRS total) are allowed to drop out, these results are not as dependable as in short-term studies. Neither the sponsor nor this reviewer made a thorough investigation about the nature of these secondary results over time.

Reviewer's Comments and Conclusions on Study 89304

There was a statistically significant lower incidence of relapse in the citalopram-treated patients compared to placebo-treated patients.

There were no statistically significant differences between the two treatment groups with respect to HAM-D Total, MADRS Total, or CGI Severity of Illness.

The results from the analyses of the effects of potential explanatory variables on time to relapse showed a statistically significant (p=.02) effect of benzodiazepine administered during the double-blind treatment period. Interaction with treatment was statistically nearly significant (p=.11) at the 10% level because the relapse rate was lower for placebo (4/39 or 10.3%) than for citalopram (10/87 or 11.5%), among the non-users of benzodiazepine.

4. **Study 89305** (Long-term)
(Original Report Vol. 1.384, Short Revised Report in Vol. 1.382)

This was a randomized, double-blind, multicenter, placebo-controlled (Phase III), parallel group study to compare the efficacy and safety (including tolerability) of two doses 20 or 40 mg per day of citalopram with placebo in the 24-week maintenance treatment of moderately to severely depressed patients who responded to citalopram in the lead-in study.

Patients were eligible to enter the study if they responded (MADRS ≤12) to citalopram or placebo treatment in clinical studies 89303 and 89306. Both of the latter studies were randomized, double-blind, placebo-controlled, parallel group, 20 or 40 mg citalopram per day, 6-week multicenter trials. Patients who responded to either dose were randomized to continue at their fixed-dose of citalopram or receive placebo during the 24-week maintenance period of Study 89305. Patients who responded to placebo in 89303 or 89306 were continued on placebo in 89305.

Baseline values were recorded at the last visit of studies 89303 or 89306 (Week 6). The primary efficacy measurement was time to relapse. Assessments were made at Weeks 4, 8, 12, 16, 20, and 24. Relapse was defined as an increase in MADRS total score to ≥22 and the clinical judgement of the investigator. Secondary efficacy measurements included: (1) MADRS Total score and (2) CGI-I and CGI-S.

Of the 207 patients enrolled in the study, 147 were citalopram-responders from studies 89303 and 89306, and constituted the primary analysis group. Forty-eight (33%) were randomized to receive 20mg/day citalopram, 57 (39%) were randomized to receive 40mg/day citalopram and 42 (29%) received placebo.

The remaining 60 of the 207 patients were placebo-responders from 89303 and 89306, were continued on placebo, and were considered to form the secondary analysis group.

Twenty-six (54%) of the 48 patients in the 20mg citalopram group completed the trial. Thirty-six (63%) of the 57 patients in the 40mg citalopram group completed the trial. Nineteen (45%) of the 42 former citalopram-responders randomized to placebo completed the trial. Twety-seven (45%) of the 60 former placebo-responders completed the trial.

A total of 28 centers with the number of patients across centers varying from 1 (centers 13, 21, 25) to 26 (center 3) were involved in this study.

The percentages of females were 60%, 63%, and 74%, respectively, in the 20mg, 40mg, and randomized-placebo group (no statistically significant imbalance between treatment groups by chi-squared test).

The median age of patients was 45 years for the 20mg group, 40 years for the 40mg group and 46.5 for the placebo group.

Race was not recorded for the two long-term European studies. The sponsor stated, "(Virtually all patients were white)."

No statistically significant differences across groups could be found in whatever the sponsor stated or presented in Tables.

Efficacy Results (Sponsor's Analyses)

The study protocol originally stated that the primary measure of interest would be the difference in relapse rates between placebo and treatment groups. However, the analysis plan was subsequently amended; the primary measure of efficacy was to be Time to Relapse, using time of withdrawal for reasons other than relapse as the censoring process.

For the ITT population, there were 4 (8.3%) relapses in the citalopram 20 mg group, 7 (12.3%) relapses in the 40 mg citalopram group, and 13 (31%) relapses in the randomized placebo group. The 20 mg citalopram and 40 mg citalopram groups each demonstrated a significantly lower relapse hazard than the placebo group (log rank test; p=0.01, and p=0.02, respectively).

Maintenance efficacy of both 20 and 40 mg doses of citalopram is shown even by the reviewer's comparison of the crude Relapse Rates (not for the period of Weeks 12-24 only).

Cumulative percentages of relapse over time (4-weekly assessments) for the two dose groups citalopram 20mg, 40 mg, and citalopram 20-40mg (combined), and placebo were provided in the Nov. 18, 1997 submission. There did not appear to be a dose effect. However, mean number of tablets per day or week for the three treatment groups was not provided by the sponsor.

The survival curve of Time to (confirmed) Relapse is attached as Figure 4.3.1.

The difference between relapse hazards of the two citalopram 20 and 40 mg groups was not statistically significant.

Secondary Analysis of Relapse: Cox Regression Analyses

The sponsor performed Cox regression analyses of the effects of the potential prognostic variables. The center variable was dichotomized into one class containing center 3 and another class containing the rest of the centers. The sponsor stated, ":this was necessary to make the model converge because of the relatively low number of patients at all centers except number 3." The age variable was also dichotomized (≥40 years against the rest).

Only the center variable had a statistically significant effect on time to relapse (p= .01). It was estimated that the relapse hazard was about three times (3.34) higher in center 3 than in other centers grouped. The sponsor stated, "However, it was not possible to analyze the possible interaction with treatment of three of the variables - center, age, and benzodiazepines - due to the low number of patients (and relapses) in the different interaction groups."

Out of the remaining 5 prognostic variables in Table 5.7 of the NDA, only for "antidepressant" there was a significant (at 10% level; p=.07) interaction with treatment. No discussion about this interaction could be found in the NDA. By request, the sponsor provided some descriptive statistics, which are presented in the next Section.

The results for the secondary efficacy variables **MADRS Total** and **CGI Severity of Illness** were, generally (not by OC analyses), statistically significant and, occassionally, marginally significant. This conclusion is based on the sponsor's reports discussing Week12 and Week-24 ITT LOCF results. The sponsor stated, "... as time goes by all patients with low MADRS scores remain in the study. Because of this, and in order to avoid bias, the secondary analysis of MADRS scores was based upon ITT LOCF."

Since relapses (considering MADRS total) are allowed to drop out, these results are not as dependable as in short-term studies. Neither the sponsor nor this reviewer made a thorough investigation about the nature of these secondary results over time.

Reviewer's Comments and Conclusions on Study 89305

The incidence of relapse was statistically significantly lower in

both the citalopram 20mg group and the 40mg group compared to that in the placebo group.

The results for the secondary efficacy variables MADRS Total and CGI Severity of Illness were also reasonably statistically significant.

There were no statistically significant differences between 20mg and 40mg citalopram treatment groups, although 20mg performed numerically better.

For "antidepressant" there was a significant (at 10% level; p=.07) interaction with treatment. Only 2 patients used additional anti-depressant, one in the placebo group and one in the citalopram group. Therefore, this finding may be unstable. This reviewer does not see any reason for this interaction in the following Table, other than "no observation" in the "citalopram 20mg and used anti-depressant" cell,

	Placebo	Citalopram 20mg	Citalopram 40mg
Anti-depressant	1/1(100%)	0/0	1/3(33.3%)
No anti-depressant	12/41(29.3%)	4/48(8.3%)	6/54(11.1%);

and asked for an explanation from the sponsor, if they can find some by tracking the analyses.

It was estimated that the relapse hazard was about three times (3.34) higher in center 3 than in other centers grouped.

III. Overall Reviewer's Comments

Study 85A provided statistical evidence for the efficacy of 20 to 80 mg/day flexible-dose citalopram in the 4-week treatment of depression. The 6-week Study 91206 provided moderate (strong with respect to HAM-D Mood Item) statistical evidence in favor of the efficacy of 40mg and 60mg per day fixed-doses of citalopram.

Study 91206 provided clear statistical evidence in favor of the efficacy of 40 and 60 mg citalopram by the protocol mentioned statistical analyses (without multiple comparison adjustments, as was desired by the protocol) of the primary efficacy variable (HAMD total). However, the clinical and statistical reviewers tried to follow the procedures they usually follow for anti-depressants; multiple comparison adjustments and robustness of results across LOCF and OC have played roles in the above conclusion.

The two long (medium)-term studies 89304 and 89305 provided clear statistical evidence with respect to relapse hazard in favor of the efficacy of citalopram in the maintenance treatment of depression.

In the two long (medium)-term studies 20mg seem to have performed numerically better than 40mg. However, in the short-term study 91206, 20mg has performed numerically worse than even 10mg.

Ninety-five percent confidence intervals for the main studies, side-by-side for endpoint (LOCF) are presented in Figures 0.2.1 to 0.2.3 for short-term and Figure 0.2.4 for long-term studies. For short-term studies, the evidence for the efficacy of citalopram was the best with respect to HAM-D depressed mood item. The HAM-D Total results also were reasonably acceptable for the efficacy of citalopram, with statistically significant results in Study 85A and for 40mg and 60mg in Study 91206. Results in other studies were numerically in favor of citalopram. With respect to CGI Severity of Illness, the only statistically significant result was for 40mg citalopram in Study 89303. Two studies produced numerically better results for placebo, than for 40mg in one and 20mg in another. These confidence intervals were not adjusted for multiple comparisons (double-checked with the sponsor's statistician by telephone).

The long-term results are acceptable. Confidence intervals were provided for the Relapse **Rate** in two groups 20 and 40 mg, although there was only one flexible dose citalopram group in Study 89304.

Comprehensive Tables of efficacy results for four ($\overline{5}$ for CGI-S) short-term studies (one page for each of the three most important efficacy variables) are attached as Tables 0.3.1 to 0.3.3. Efficacy results (Relapse) for the two long-term studies are attached as Table 0.3.4.

From the pairwise comparisons provided (Nov. 7, 1997 submission), we see that in Study 91206, 40mg and 60mg results were, more often than not, statistically significantly or marginally significantly superior to 20mg. In this study, 20mg results were poor.

Consistency Across Sites

In Study 85A, the results across the three sites were reasonably consistent.

In Study 91206, out of the twelve centers, a few centers had

placebo results better than the citalopram (40 and 60 mg doses) results. For 20 mg, less than 50% centers had citalopram (20mg) results better than the placebo results (overall results also were the poorest for this dose). It does not seem that the overall efficacy of the higher doses was driven by a few big centers.

For the long-term studies, the sponsor provided some by-center results (95% confidence intervals) for MADRS total in the 12-12-97 submission and concluded, "From the results of ANOVA, it was found that none of the treatment-by-center interactions were statistically significant in both long-term studies". Graphically, there were some centers where placebo did better than citalopram.

For Study 89305, it was estimated that the relapse hazard was about three times (3.34) higher in center 3 than in other centers grouped.

Subgroup Analyses (Race, Gender, Age, Baseline Severity)

The sponsor stated, "... we analyzed the interaction of treatment by four different covariates-- age, race, gender, and baseline severity. ... The only statistically significant interaction term in the analyses aforementioned is age-by-treatment in Study 85A."

The combined results for four short-term studies showed that Age, Race, Gender, and Baseline Severity were statistically significant covariates with respect to at least one (Age in all, Gender and CGI Severity in two) of HAM-D Total, HAM-D Item 1, and CGI Severity.

The sponsor concluded (submission of Nov.18, 1997), "In Study 85A, female patients tend to do better numerically or statistically than male patients in all three efficacy variables. In Study 91206, this tendency was not observed. For two studies pooled together, females had better improvement than males in HAM-D total score and Mood score, but not in CGI-S score. For the other three covariates, there is no consistent trend of one subgroup doing better than any other subgroup."

There was a significant age-by-treatment interaction with respect to HAM-D total and HAM-D Item 1, in Study 85A because for patients greater than 60 years old, placebo (n=4) did better than citalopram (n=7). However, with such a small number of patients in this age group, we cannot confirm that citalopram is inefficacious or worse than placebo in this age group. Combined

results for four short term studies did not show this interaction.

Benzodiazepine studied in one **long-term** Study 89304 and "**antidepressant**" studied in the other long-term Study 89305 indicated some interaction with treatment (interaction p-values .11 and .07 respectively; details in "Reviewer's Comments" for each study).

IV: Overall Conclusion

Statistical evidence in favor the efficacy of citalopram in the short-term treatment of depression has been provided clearly by one and moderately (also clearly if multiple comparison adjustments are not applied, as desired by the protocol) by another of the short-term studies.

The two long (medium)-term studies provided clear statistical evidence with respect to relapse hazard in favor of the efficacy of citalopram in the maintenance treatment of depression.

In the two long (medium)-term studies 20mg seem to have performed numerically better than 40mg. However, in the short-term study 91206, 20mg has performed numerically worse than even 10mg.

A few marginal interactions (each one in a separate study) have been discussed above.

/S/ 2-26-98

Japobrata Choudhury, Ph.D.
Mathematical Statistician

Concur: Dr. Sahlroot /S/ 3/2/98

Dr. Chi /S/ 3/9/98

CC:
Archival NDA 20-822

HFD-120/Dr. Leber
HFD-120/Dr. Laughren
HFD-120/Dr. Dubitsky
HFD-120/Mr. Purvis
HFD-120/Mr. David

HFD-344/Dr. Barton
HFD-710/Dr. Chi
HFD-710/Dr. Sahlroot
HFD-710/Dr. Choudhury
HFD-710/Chron
J.Choudhury:x45582:DB I: 02/26/98

This review consists of 28 pages of text and 41 pages of Tables, Figures, etc.

APPEARS THIS WAY
ON ORIGINAL