CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

18-936/SE5-064

STATISTICAL REVIEW(S)

STATISTICAL REVIEW

NDA:	**18-936**
SPONSOR:	**Lilly**
DRUG:	**Fluoxetine (Prozac)**
INDICATION:	**Pediatric OCD and Depression**
MATERIAL SUBMITTED:	**Statistical Review**
DATE SUBMITTED:	**10/10/2000 (Acute Treatment Period)**
	1/15/2001 (Relapse Prevention)
MEDICAL OFFICER:	**Andy Mosholder, M.D.**

I. BACKGROUND

The sponsor submitted a supplemental NDA that was used to fulfill the requirement of pediatric exclusivity written request and to support depression and OCD (Obsessive-compulsive Disorder) claim in the pediatric population.

Two studies were submitted to support the **pediatric depression** claim :

Study B1Y-MC-X065 : " Fluoxetine Versus Placebo in the Acute Treatment of Major Depressive Disorder in Children and Adolescents", and

Study B1Y-MC-HCJE : " A Multicenter, Double-blind, Randomized, Parallel-group Placebo-controlled Study".

One study was used to support the pediatric **OCD** claim :

Study B1Y-MC-HCJW : "Fluoxetine Versus Placebo in the Treatment of Children and Adolescents with Obsessive-Compulsive Disorder".

II. Study B1Y-MC-X065 (Pediatric Depression)

This study was conducted at the University of Texas Southwestern Medical Center at Dallas from April 1991 to February 1995. The original protocol for this study was a NIMH grant proposal submitted by Dr. Emslie. The inclusion criteria consists of outpatients with non-psychotic, major depressive disorder (MDD), single and recurrent episodes according to DSM-III-R (Diagnostic and Statistical Manual of Mental Disorders, third edition-revised), who were aged 8 to 18 years and normal intelligence, and who were willing and able to sign informed consent (parents and patients). Diagnosis of MDD also depended on whether patients having a Children's Depression Rating Scale-Revised (CDRS-R) total score >40 at study entry. The diagnosis was finally decided at a consensus meeting of the clinical investigators.

II.1 Study design

This was a single-center, double-blind, randomized, parallel-group study. Two phases were described in the original protocol : 3-week diagnostic evaluation phase and an acute treatment phase which includes a 1 to 2 week placebo lead-in period followed by an 8-week acute treatment period in which patients were randomized to either 20mg/day fluoxetine or placebo. The sponsor was mainly interested in the second phase of the study, so the sponsor only collected data from the second phase of Dr. Emslie's study. The sponsor also collected the data indicating the final consensus diagnosis, but not the actual scales and evaluations that were used to achieved the final consensus. Since the site did not enter all variables into the database, the sponsor designed electronic case report forms and asked the site personnel to enter additional information based on study files for each patients, i.e. source documents. These retrospective collected data are mainly for safety information which includes non-solicited adverse events, pill counts, laboratory data, concomitant medications, ECGs, vital signs, and some inclusion/exclusion criteria.

In this study, the sponsor only focused on the second period, so they treated the single-blind placebo lead-in period as the study period I and the double-blind, acute treatment period as study period II (note: patients were randomized at Visit 2 which corresponds to the baseline visit of the double blind, 8-week, acute treatment period). This reviewer will follow the sponsor's visit naming rule throughout this statistical review.

II.2 Objectives

In the original protocol, to compare the efficacy of fluoxetine and placebo in the treatment of MDD in children and adolescents was **one of the objectives**. The primary endpoint for this comparison was the proportion of **completing subjects** in each group who recover which was defined as below 28 on the CDRS-R and a CGI of 1 or 2.

In the sponsor's report, they specified the primary objective as comparing fluoxetine versus placebo in the treatment of children and adolescents diagnosed with MDD as measured by response rates (defined as at least 30% decrease from baseline in the CDRS-R total score) after up to 8 weeks. The sponsor provided the rationale of choosing 30% reduction in CDRS-R score as clinically significant: At study entry, patients were required to have a score of >40 on the CDRS-R in order to be enrolled (i.e. minimum score associated with active depression). Remission of depressive symptoms is defined as achieving a score of 28 or less on the CDRS-R. The difference between the minimum entry criterion for depression and remission is 30%.

II.3 Efficacy Endpoint

As indicated in the "Objective section", the protocol specified endpoint is "recovery" defined by the last measurement of CDRS-R total score ≤ 28 or CGI-improvement score of 1 or 2 from Visits 3 through 10. However, the sponsor indicated in the study report that they defined objectives for analysis of the acute treatment phase prior to any reanalysis of the unblinded data. Instead of using both CDRS-R and Clinical Global Impressions (CGI) improvement scores to define the primary endpoint, the sponsor defined response based on a 30% reduction in CDRS-R total scores from baseline (last measure from Visits 1 and 2) to endpoint (last measure from Visit 3 to 10) as the primary efficacy variable.

Other comparisons based on mean change in CDRS-R total and subtotal scores from baseline to endpoint, remission (defined as a CDRS-R total endpoint score of ≤ 28), response rate based on the CGI-improvement scores (score 1 or 2 was the responder) and other scales, etc, were treated as the secondary endpoints in the sponsor's report.

II.4 Number of Subjects and Analysis Plan

Based on Dr. Emslie's original protocol, the sample size was calculated based on the percent recovered. Assuming 40% recovery rate in placebo group and 70% in the fluoxetine group, 40 patients were obtained for each treatment group to assure 80% power to detect the treatment difference with 0.05 type I error rate.

In Dr. Emslie's original protocol, the primary endpoint (proportion of recovery) would be tested by Chi-square based on completing subject. But in this review, the Intent-to-treat (ITT) population would be used for all analyses. Only patients who had baseline and post-baseline measures were included in the efficacy analyses.

In the sponsor's statistical analysis plan (prior to unblinding of the sponsor's in-house data), the following specifications were provided: All tests were performed at a 2-sided, 0.05 significance level; for missing items in a scale, the total or subtotal was treated as missing. The primary efficacy analysis of the percent response (at least 30% decreased in CDRS-R score from baseline) was the Fisher's exact test (which was a post hoc specification). The sponsor also performed additional analyses to augment the results from the primary analysis : a 50% reduction from baseline for ITT populations and a 30% and 50% reduction from baseline for patients who finished at least 4 weeks.

The sponsor also performed subgroup analysis on the primary efficacy endpoint. The subgroups include age strata (8 to <13 and 13 to ≤ 18) and gender. Breslow-Day test was used to test the homogeneity of odds ratio results to evaluate the between-strata difference with respect to treatment effect. .

Categorical endpoints such as proportion of patients met the remission criteria and CGI-improvement response were also compared between treatment groups using a Fisher's exact test.

The mean change in CDRS-R total score from baseline to endpoint (last available measure from Visits 3 through 10) was compared across treatment groups using an ANOVA with treatment in the model. The following secondary variables were analyzed similarly :

- CDRS-R Mood Subtotal – sum of items 8, 11, 14, 15
- CDRS-R Somatic Subtotal – sum of items 4, 5, 6, 7, 16, 17
- CDRS-R Subjective Subtotal - sum of items 9, 10, 12, 13
- CDRS-R Behavior Subtotal – sum of items 1, 2, 3
- CGI-severity
- BPRS-C total
- BDI total
- CDI total

The longitudinal analyses were also performed for the change of CDRS-R total scores from baseline. The methods include fitting ANOVA (with treatment in the model) by visit based on observed case or LOCF method and repeated measured ANOVA. In the repeated measured model, the dependent variables were the baseline and all the post-baseline CDRS-R total scores. The model used an unstructured covariance matrix including treatment, visit, and an interaction term of treatment by visit in the model. A single degree-of-freedom contrast of comparing the difference between baseline to Visit 10 scores across treatment groups was evaluated. Inference from the repeated measure analysis was based on the restricted maximum likelihood method and from its approximate F-test.

II.5 The Sponsor's Documentation of the Study Conduct and Retrospective Data Collection

In the study report, the sponsor had documented and raised some issues about the study conduct. Some of the issues were noted in the following (For more detailed documentation, please refer to section 9.4.6: Treatment Blinding) :

- From April 1991 to August 1993, the pharmacy at the Children's Medical Center at Dallas prepared and provided blinded study drug medication for the study. Fifty-four patients (25 fluoxetine-treated patients and 29 placebo-treated patients) received medication in this manner. But from September 1993 to February 1995, Lilly supplied blinded clinical trial material for this study. Forty-two patients (23 fluoxetine-treated and 19 placebo-treated patients) received medication in this manner. The sponsor did not find significant treatment difference from both sources of treatment preparation. Therefore, they concluded that the source of treatment preparation should not bias the study result.
- A study site nurse served as the liaison between the clinical site and the pharmacy during the acute treatment phase. She involved with rating of patients during the

diagnostic evaluation and had access to the randomization list. This nurse signed an affirmation statement that said she did not complete any post-randomization rating except subject 2013 at visit T2 and subject 2014 at visit T1. The sponsor believes that her evaluation on these 2 patients did not compromise the study result.

- An affirmation statement was sent out by the South western Medical center with statement as "all staff including Dr. Emslie were not unblind to patient therapy code until June 1995, when all patients completed the study and all databases were cleaned and checked" to confirm that the investigator remained blinded during the study.
- Patients who completed the study were referred to the physicians for follow-up care. Some of these physicians were involved in the study. This study conduct had inadvertently unblinded less than 10 patients to the study physicians prior to the completion of the entire study. Due to the limited number of occurrence, the sponsor does not believe this compromised the overall results.

Since the data was not monitored by the sponsor while the trial was conducted (1991-1995), a series of measures were taken to ensure the integrity of the data (For more detailed documentation, please refer to section 9.6 : Data Quality Assurance) :

- A detailed plan to maintain the study blind at the patient level was developed;
- An extensive audit of the source documents and study files was conducted;
- Affirmation statements from the Investigator and Study Coordinators were obtained;
- An audit trail for the study database was initiated and maintained;
- 100% source data verification and 100% quality review of all data points for every patient at every visit (the sponsor's database) during the acute treatment period was conducted;
- A statistical analysis plan was developed prior to reanalyzing any unblinded data;
- A data validation plan was developed to document the data collection procedure.

II.6 Sponsor's Results

A total of 108 patients were screened for entering the acute treatment period. 12 of these patients did not meet the entry criteria or decided not to participate, so only 96 patients were randomized at Visit 2 (48 patients received fluoxetine and 48 received placebo). Among these patients, 33 (69%) fluoxetine patients and 25 (52%) patients completed the study. The main reason for early discontinuation is lack of efficacy : 10.4% of fluoxetine patients and 39.6% of placebo patients (Table A.I.1). Of the 96 patients randomized, only one patient (2207) was not included in the efficacy analysis due to missing a post-baseline CDRS-R assessment.

Table A.I.1 Summary of Reasons for Discontinuation
All Randomized Patients

Primary Reason for Discontinuation	Flx (n=48)		Placebo (n=48)		Total (n=96)	
	n	(%)	n	(%)	n	(%)
Complete Period IV	33	(68.8)	25	(52.1)	58	(60.4)
Adverse Event	5	(10.4)	0		5	(5.2)
Lack of Efficacy	6	(12.5)	19	(39.6)	25	(26.0)
Patient decision	0		2	(4.2)	2	(2.1)
Protocol Requirement	3	(6.3)	2	(4.2)	5	(5.2)
Physician Decision	1	(2.1)	0		1	(1.0)

The treatment assignment was balanced with respect to the number of adolescents (n=48) and number of children (n=48) with fluoxetine and placebo assigned equally in each age group. Among the randomized patients, 44 (46%) were females and 52 (54%) were males. The majority of these patients were Caucasian (72.9% of the fluoxetine group and 85.4% of the placebo group). The distribution of the baseline patient characteristics such as height, weight, socioeconomic status and family structure, appear to be comparable between treatment groups (Table A.I.2).

Most psychiatric history were comparable between treatment groups, except some psychiatric diagnoses: more comorbid ADHD in fluoxetine group (29%) than that in placebo (18%), two times of comorbid anxiety in fluoxetine group (54.2%) as compared to placebo (27.1%) and more dysthymia in fluoxetine group (42.7%) than that in placebo (29.2%). The difference in the percentage of comorbid anxiety achieved statistical significance based on the Fisher's exact test (p-value=0.012). The most common comorbid disorders included anxiety (41%), dysthymia (35%), oppositional and conduct disorder (29%) and ADHD (Attention Deficit Hyperactivity) (24%). More than half of these patients had positive first-degree family, Axis I, disorder, however it was balanced between treatment groups (52.1% of fluoxetine versus 56.3% of placebo).

Most of the patients did not receive medication for their current MDD episode (60%). Among 30 patients who had received previous medication, most patients used tricyclic antidepressants (14%) and psychotherapy (13%).

With respect to the baseline psychiatric evaluation on CDRS-R, CGI-severity and BPRS-C (Brief Psychiatric Rating Scale) total scores, there was no baseline disparity found in these scores (Table A.I.3).

Table A.I.2 Baseline Patient Characteristics
All Randomized Patients

Variables		Fluoxetine (n=48)	Placebo (n=48)	Total (n=96)
Age	Mean	12.67	13 00	12.84
	Median	13.00	12.98	12.98
	SD	2.73	2.78	2.75
	Min.	7.56	7.16	7.16
	Max.	17.84	17.80	17.84
Age category				
	8 - <13 yrs.	24(50.0)	24 (50.0)	48 (50.0)
	13 - <18 yrs.	24(50 0)	24 (50.0)	48 (50.0)
Gender	Female	22 (45.8)	22 (45.8)	44 (45.8)
	Male	26 (54.2)	26 (54.2)	52 (54.2)
Race	White	35 (72 9)	41 (85 4)	76 (79.2)
	Black	4 (8.3)	4 (8.3)	8 (8 3)
	Hispanic	8 (16.7)	2 (4.2)	10 (10.4)
	Other	1 (2.1)	1 (2.1)	2 (2.1)
Weight (kg)				
	N	44	46	90
	Mean	54.07	50.83	52.41
	Median	51.00	49.00	50.00
	SD	21.82	15.48	18.81
	Min.	23.00	26.00	23.00
	Max.	120.00	99.00	120.00
Height (cm)				
	N	37	39	76
	Mean	149.93	153.93	151.99
	Median	149.00	153.00	153.00
	SD	20.58	12.59	16.96
	Min.	66.00	137.00	66.00
	Max.	180.00	180.00	180.00
CDRS-R Total				
	N	48	48	96
	Mean	58.9	57.5	58.2
	Median	57.5	54.5	57.0
	SD	10.4	10.3	10 3
	Min.			
	Max.			

Table A.I.3 Baseline Psychiatric Evaluation
All Randomized Patients

Variables	Fluoxetine (n=48) n(%)	Placebo (n=48) n(%)	Total (n=96) n(%)
Comorbid Ax 1 Diag ADHD :			
No	34(70.8)	39(81.3)	73(76.0)
Yes	14(29.2)	9(18.8)	23(24.0)
Comorbid Ax 1 Diag Alcohol abuse			
No	48(100.0)	47(97.9)	95(99.0)
Yes	0	1(2.1)	1 (1.0)
Comorbid Ax 1 Diag Anxiety			
No	22 (45.8)	35 (72.9)	57 (59.4)
Yes	26 (54 2)	13 (27.1)	39 (40.6)
Comorbid Ax 1 Diag Dysthymia			
No	28(58.3)	34(70.8)	62(64.6)
Yes	20(41.7)	14(29.2)	34(35.4)
Comorbid Ax 1 Diag Func. Enuresis			
No	47(97.9)	48(100)	95(99 0)
Yes	1(2.1)	0	1(1.0)
Comorbid Ax 1 Diag Obsessive Compulsive			
No	47(97.9)	47(97.9)	94(97.9)
Yes	1(2.1)	1(2.1)	2(2.1)

In the responder analysis based on Dr. Emslie's original primary efficacy variable (i.e. recovery rate, see Table A.I.4), the result was not significant (29% for Fluoxetine-treated group and 19% for placebo, p=0.339). The analysis based on remission was also not significant (p=0.238). But the result was significant based on the percent responder defined by CGI-improvement scores (p=0.040). Note that in Dr. Emslie's '97 paper[1], the percent responders defined by CGI-improvement was presented as the primary result (p=0.02 based on Chi-sqaure test was reported in the paper instead of 0.04 based on 2-sided Fisher's exact test). In this review, the p-values for categorical variable analysis were reported based on 2-sided Fisher's exact test, unless otherwise specified. This reviewer did not find any change of the result if a Chi-square test statistics (protocol specified) would be used instead.

[1] G.J. Emslie; A. J. Rush, et.al, A Double-blind, Randomized, Placebo-Controlled Trial of Fluoxetine in Children and Adolescents with Depression, Arch Gen Psychiatry, 1997;54:1031-1037.

In the sponsor's post hoc analysis of 30% reduction from baseline based on CDRS-R scores, the sponsor obtained a response rate of 58.3% for the fluoxetine and 31.9% for the placebo group (p=0.013, Table A.I.5). The same comparison for all patients who completed at least 4 weeks of treatment was also significant (p =0.031). However, the comparison based on 50% reduction from baseline on CDRS-R scores was not significant (p=0.467). The sponsor also evaluated other cut-off points (from 0% to 60%) to define the percent reduction. The largest treatment difference was found when 20% or 30% cut-off points was chosen (note: p =0.007 for 20% reduction).

Table A.I.4 Efficacy Endpoints :
% Responders based on Remission, Recovery and CGI-improvement Scores

Response	Therapy	N	No		Yes		p-value (2-sided Fisher's exact)
			N	%	N	%	
Remission	Flx 20 mg	48	33	68.6	15	31.3	0.238
	Placebo	47	38	80.9	9	19.1	
Recovery	Flx 20 mg	48	34	70.8	14	29.2	0.339
	Placebo	47	38	80.9	9	19.1	
CGI-improvement	Flx 20 mg	48	21	43.8	27	56.3	0.040
	Placebo	47	31	66.0	16	34.0	

Table A.I.5 CDRS-R Total Score
Number of patients Meeting Criteria for Response
(30% reduction from baseline)

Therapy	N	No		Yes		p-value (2-sided Fisher's exact)
		N	%	N	%	
Flx 20 mg	48	20	41.7	28	58.3	0.013
Placebo	47	32	61.8	15	31.9	

In a secondary efficacy analysis of the mean change in CDRS-R total scores from baseline (based on LOCF), the reduction of CDRS-R total score is significant lower in fluoxetine group (-20.2) than the placebo (-10.5) (p-value=0.002). In a similar analysis based on observed case only, the difference was not statistical significant (p-value=0.106, -24.2 for fluoxetine group and –18.2 for placebo).

In the subgroup analysis of the percent of patients with 30% reduction from baseline in CDRS-R total score, fluoxetine showed a more favorable result over placebo consistently across different subgroups (Table A.I.6). The consistency was also demonstrated when change in CDRS-R total score from baseline was analyzed (see sponsor's table 11.20).

Table A.I.6 CDRS-R Total Score by Subgroup
Number of patients Meeting Criteria for Response
(30% reduction from baseline)

Subgroup	Therapy	N	Response N	Response %	p-value (2-sided Fisher's exact)
Age					
8 to <13	Flx 20 mg	24	15	63.0	.148
	Placebo	23	9	39.0	
13 to ≤18	Flx 20 mg	24	13	54.0	.075
	Placebo	24	6	25.0	
Gender					
Female	Flx 20 mg	22	12	55.0	0.124
	Placebo	21	6	29.0	
Male	Flx 20 mg	26	16	62.0	0.095
	Placebo	26	9	35.0	

II.7 Reviewer's Evaluation and Comments

In a letter to the sponsor dated April 12, 1999, the agency had the following comments related to the sponsor proposed primary endpoint : "As discussed at our March 24, 1998 meeting, we do not consider your designated primary outcome measure, i.e. proportion of patients achieving a ≥ 30% reduction from baseline to endpoint on the CDRS-R to be the best choice for a primary outcome, and we will consider other measures as well in our overall judgement regarding the outcome of these trials". This statistical review will start with the evaluation of the protocol specified primary efficacy endpoint first. But other endpoints will be evaluated as well.

This reviewer confirmed the protocol specified primary analysis result that showed **non-significant treatment effect** in a direction favoring fluoxetine (p=0.339). The responder analysis based on CGI-improvement scale (as the primary endpoint appeared in Dr. Emslie's '97 paper) was significant (p=0.040). In addition, this reviewer confirmed the sponsor's post hoc analysis (30% reduction in CDRS-R total score) result in favor of fluoxetine (p=0.013).

Since the sponsor's analysis did not take consideration of the stratified randomization scheme, this reviewer performed the analysis for all the responder analyses (recovery,

remission, responders based on 30% reduction in CDRS-R total score or based on CGI-improvement scale) using the Cochran Mantel Haenszel Statistics, controlling for strata (all combination of age group and gender). These stratified analyses did not changed the results much, e.g. the treatment effect for recovery, remission rates remained non-significant, while it remained significant for the analysis based on 30% reduction in CDRS-R total score or CGI-improvement score. Similarly, the result was significant by fitting ANOVA model for change in CDRS-R total score from baseline including treatment and strata (p=0.002).

This reviewer noticed that if the cut-off point moved up to 40% or 50% reduction in CDRS-R total scores, the results were not significant based on 2-sided Fisher's exact test (p=0.089 for 40% reduction and p=0.467 for 50% reduction). The largest treatment effect was found when 20% and 30% were chosen as the cut-off points (p=0.007 and 0.031 for 20% and 30 % cut-off point, respectively).

According to the sponsor, an imbalance treatment assignment was found for the baseline comorbid anxiety : more fluoxetine-treated patients had baseline comorbid anxiety as compared to placebo patients. Medical reviewer, Dr. Mosholder pointed out the potential impact of the study result from such imbalance treatment assignment. This reviewer then performed a subgroup analysis using the sponsor's post hoc endpoint (30% reduction in CDRS-R score) based on baseline comorbid anxiety (Table A.II.1) status. The result showed that for patients with baseline comorbid anxiety (n=39), the treatment effect was highly significant in favor of fluoxetine (p=0.008), while for patients without baseline comorbid anxiety (n=56), no statistical significant treatment benefit was found (p=0.278).

For responder analysis based on CGI-improvement score (Primary endpoint appeared on Dr. Emslie's '97 paper), treatment effect for patients with baseline comorbid anxiety was marginally significant (p=0.051), while for patients without anxiety comorbidity was not significant (p=0.278).

Similar results were found when the change of CDRS-R total score from baseline was analyzed by baseline comorbid anxiety subgroup (p=0.0016 and p=0.1081 for patients with baseline comorbid anxiety and without baseline comorbid anxiety, respectively). Nevertheless, the significant treatment effect was not changed when the ANOVA model was fitted for the change of CDRS-R total score including treatment and baseline comorbidity as two factors. Since more patients were assigned to the fluoxetine group for patients with baseline comorbid anxiety, the highly significant result in this subgroup may contribute to the overall significant result. A further evaluation may be needed to address the concern of the possible disparity of the treatment effect between the baseline comorbid anxiety status.

Table A.II.1 CDRS-R Total Score by Baseline Comorbid Anxiety
Number of patients Meeting Criteria for Response
(30% reduction from baseline)

Subgroup	Therapy	N	Response N %	p-value (2-sided Fisher's exact)
Comorbid Anxiety				
No	Flx 20 mg	22	12 54.6	.278
	Placebo	34	13 38.2	
Yes	Flx 20 mg	26	16 61.5	.008
	Placebo	13	2 15.4	

Although the sponsor had extensive documentation of the study conduct and retrospective data collection, these documents can not substitute a prospective designated plan. The sponsor tried to assure that the study result was not compromised by the study conduct. But this reviewer believes that some of the impact was un-measurable, for example, the study unblinded coordinator who was also involved in the study operation. She may not directly rate the patients' endpoint scale, but indirectly, she may have impact on the study result. Particularly, this is a single center study and she may be the only one that coordinated the entire study.

III. Study B1Y-MC-HCJE (Pediatric Depression)

This study was conducted from April 27, 1998 to July 21, 2000 (with the interim report cut-off date December 16, 1999). Note : the interim analysis database was locked on February 18, 2000 and the final database (for relapse prevention) was locked on August 17, 2000; the statistical analysis plan was signed off on February 7, 2000.

The final report for acute treatment period and relapse prevention period were dated August 6, 2000 and January 8, 2001, respectively. The interim report was submitted in October 10, 2000 as the first part of the NDA submission for acute treatment period. The final report was submitted on January 15, 2000 as a separate part of the NDA submission for relapse prevention.

Patients who were aged 8 to <18 years, who had a primary psychiatric diagnosis of nonpsychotic major depressive disorder (single or recurrent) as determined by DSM-IV criteria, who had depressive symptoms of at least moderate severity (defined by a CDRS-R total score > 40 and a CGI-severity rating of moderate or greater), who were able to swallow whole medication, who had no clinical significant laboratory or ECG findings,

who could communicate intelligibly with the investigator and study coordinator and who (and their parents) agreed to keep appointments, were included in this study.

III.1 Study Design

This was a multicenter, randomized, double-blind, two-arm, parallel study. Patients were randomized at Visit 4 based on their gender and age category (i.e. stratified randomization) across investigative sites. The study had two phases of interest, **an acute treatment phase and a relapse prevention phase**. There were a total of six periods :

Period I : The diagnostic evaluation phase in which three separate diagnostic interviews were conducted in three visits (i.e. beginning with visit 1 and end with Visit 3 with 6-9 days between visits). Patients who met the inclusion/exclusion criteria at all three visits, and had a CDRS-R score of > 40 at Visit 3, entered Period II.

Period II : A 1-week placebo wash-out phase (i.e. visits 3-4). Patients who responded to placebo (defined as a ≥30% reduction in CDRS-R score from Visit 3 to Visit 4 or a CGI-improvement score of 1 or 2 at Visit 4 as compared to Visit 1) was discontinued. Patients who did not respond to placebo entered Period III, a 1-week adaptation phase.

Period III : A 1-week adaptation phase(visits 4-5). Patients who met the inclusion/exclusion criteria at Visit 4 were randomized (stratified by gender and age category : child versus adolescent, across investigative sites).

Patients who were randomized to receive fluoxetine started 10 mg for a week. Patients who adapted well continued into study Period IV, otherwise, they were discontinued.

Period IV : An 8-week fixed-dose phase (Visits 5-10). Visits were scheduled weekly (6-9 days) through Visit 7 and biweekly (13-16 days) thereafter. Patients were titrated to 20 mg at Visit 5. If they cannot tolerate 20 mg, the dose was reduced to 10 mg at Visits 6 –9.

Period V : A 10-week titration phase (Visits 10-15). Fluoxetine or placebo responders stayed with their assigned treatment (i.e. fluoxetine 20 mg or placebo). The fluoxetine non-responders were re-randomized to either remain on fluoxetine 20 mg or to receive fluoxetine 40 mg with an option to titrate to 60 mg. The placebo non-responders stayed on placebo.

Period VI : A 32-week double-blind relapse prevention phase (Visits 15 to 26). Patients who received fluoxetine 20-60 mg/day, with a CDRS-R score of ≤28 at Visit 15 were eligible to participate in the relapse prevention phase. Patients were re-randomized to placebo or to continue their current treatment as of Visit 15. Both Investigators and patients were blinded to the actual timing of the switch (Visit S) to the new treatment. The visits were scheduled biweekly for the

first 3 months and monthly (26 to 32 days) thereafter.

The review of the efficacy analysis contains two phases :

- **Acute treatment phase and sub-chronic treatment phase :**
  Acute treatment phase included 9-week treatment period which refers to the period from baseline (Visit 4) to period III through period IV (Visit 5 to 10); sub-chronic treatment phase included 19-week treatment period which refers to the period from baseline (Visit 4) to period III through period V (Visit 5 to 15);

- **Relapse prevention phase :**
  Study period VI was the relapse prevention phase. The data in this phase was presented in a separate clinical study report.

Since only fluoxetine non-responders were rerandomized in Period V and Period V was not the period that the primary efficacy endpoint was evaluated, this review will only cover the sub-chronic treatment phase briefly in this review.

III.2 Objectives

The primary objective of this study was to compare the response rate (defined as the percentage of patients who had at least 30% reduction of CDRS-R (Children's Depression Rating Scale-Revised, CDRS-R) from baseline to endpoint (Visit 5 to 10) between fluoxetine 20 mg and placebo in the treatment of children and adolescents with DSM-IV major depression.

III.3 Efficacy Endpoints

III.3.1 Acute and Sub-chronic Treatment Phases (Refers to Period III through Period IV or Period III through V, respectively)

Since the agency did not concur with the sponsor's 30% reduction on CDRS-R total score endpoint (see section II.3 Efficacy endpoint of study X065), this review will first review the protocol specified primary efficacy endpoint and will evaluate other endpoints as well.

The primary efficacy endpoint was the percent response defined as at least 30% reduction on CDRS-R from baseline (Visit 4) to endpoint **(last measurement from Visit 6 through 10)** (Note : the sponsor referred this 9-week period as the acute treatment phase). This analysis included patients treated at least one week with 20 mg (2 weeks total).

CDRS-R is a clinician-rated instrument designed to measure the presence and severity of depression in children. It consists of 17 items scored on a 1 to 5 or 1 to 7-point scales. A

rating of 1 indicates normal functioning. Total scores ranged from 17 to 113. In general, scores below 20 indicate absence of depression, scores of 20 to 30 indicate borderline depression, and scores of 40 to 60 indicate moderate depression.

The response rate from baseline to the last measurement between Visits 6 to 15 treated as the secondary endpoint (**Note : the sponsor referred this 19-week period as the sub-chronic treatment phase; it was stated in the protocol, but not in the report**). The dose for the fluoxetine non-responders can be titrated up to 60 mg/day).

The change in CDRS-R score from baseline (Visit 4) to study Periods III through IV endpoint **(last measurement from Visits 5 to 10)** and from baseline to study Period III through V **(last measurement from Visits 5 through 15)** were also compared between treatment groups. Similar change score analyses were also performed for CGI-severity, CGI-improvement, K-SADS, MADRS, CDI, BDI, HAMA.

The remission rates based on CDRS-R scores (defined as a patient who has an endpoint CDRS-R score ≤ 28) at two endpoints (Period IV and Periods IV through V endpoints) were also analyzed as the secondary endpoints. Analysis of remission included only those patients treated at least 1 week with 20 mg.

III.3.2 Relapse Prevention Phase (Study Period VI)

Time to relapse was compared between fluoxetine (10-60 mg) and placebo. Relapse during this phase was defined as a one-time CDRS-R score of >40 in the presence of a history of 2 weeks of clinical deterioration as determined by patient report, parent report, or clinical history. If the physician felt patients had experienced relapse sufficient to discontinue the patient from the study, but he/she had not met the relapse criteria, the patient was considered as a relapser for the **primary analysis of relapse**.

Patients who met the protocol-defined criteria were also analyzed as a **secondary analysis** of relapse.

Change in CDRS-R, CGI-Severity, CGI-Improvement, K-SADS, MADRS, CDI, and BDI scores from baseline (Visit S) to each visit during Phase VI was treated as the secondary efficacy endpoints.

III.4 Number of Subjects and Analysis Plan

The sample size of 110 patients per treatment arm was obtained to detect 20 % difference in percent response between fluoxetine 20 mg (assume 70%) and placebo (assume 50%) with 80% power, assuming two-sided 0.05 significance level and no more than 7 patients dropped out in each arm.

III.4.1 Acute and sub-chronic Treatment Phases (Refers to Period III through Period IV or Period III through V, respectively)

Analyses were done on an intent-to-treat basis unless otherwise specified. An intent-to-treat analysis is an analysis by grouping based on random allocation, regardless of what patients actually took or whether the patients followed the protocol. Patients who had their dose reduced to 10 mg per day were included as if they remained on 20 mg per day.

All randomized patients who had at least one post randomization visit were included in the efficacy analysis for Periods III through V. All patients with a Study Period VI baseline visit (Visit S) and at least one visit beyond baseline were included in the relapse prevention analysis.

Investigators who enrolled no patients into one or more treatment groups were pooled in the analysis when assessing treatment-by-investigator interactions.

General analysis strategies were proposed for the efficacy analysis in the protocol :

- Fisher's exact test was used to compare percent response. Note that Pearson's chi-square test was specified in the original protocol. The sponsor indicated that the change was made in the statistical analysis plan dated February 7, 2000. The logistic regression models with treatment, investigator, gender, age group and treatment by investigator interaction in the model was treated as the secondary analysis.

- ANOVA with treatment in the model will be used to compare change scores or endpoint scores between treatments. In the protocol, it specified that analyses will be performed on both the original and the rank-transformed data. The analysis of the original data will be considered as primary unless there is evidence of non-normality. ANCOVA Model with treatment, investigator, gender, age group and treatment by investigator interaction in the model was treated as the secondary analysis.

III.4.2 Relapse Prevention Phase (Study Period VI)

Time to relapse was analyzed based on log-rank test, along with Kaplan-Meier survival estimates.

Change in CDRS-R, CGI-Severity, CGI-Improvement, K-SADS, MADRS, CDI, and BDI scores from baseline (Visit S) to each visit during Phase VI was summarized by

1) including observed case only at each visit, and

2) including all patients with measures beyond Visit 17 using LOCF approach.

The same analysis methods used in acute and subacute treatment phases were used for the analysis of these secondary endpoints.

Subgroup analyses were performed for efficacy analysis of CDRS-R and CGI, by age group, gender and children with a family history of depression.

III.5 Interim Analysis

Only one interim analysis, which was the final analysis of phases I through V (after all patients completed study Period V, i.e.Visit 15), was performed (this was submitted in 10/10/2000 as the first part of the NDA submission). The protocol indicated that the results from the fixed-dose acute treatment unblinded at the treatment level may be presented outside the company including regulatory agencies. However, the data monitoring board (DMB) assigned to this study decided that the treatment results unblinded at the treatment level would only be presented to regulatory agencies.

The study was not stopped even if a significant efficacy result was observed, so no adjustment was required.

To minimize the bias for the relapse prevention therapy phase, no one at the study site was unblinded at the patient level. The Lilly clinical research administrator and clinical research physician communicating with the sites also remained blinded to treatment assignment.

III.6 Sponsor's Results

Two hundred nineteen patients were randomized to treatment at Visit 4 (119 received fluoxetine and 100 received placebo). One hundred fifty-eight (72%) patients completed acute treatment including 90 (83%) fluoxetine-treated and 68 (62%) placebo-treated patients. The primary reasons for discontinuation during the acute treatment phase were shown on Table B.I.1.

Seventy five patients (34%) completed sub-chronic treatment with 40 (37%) fluoxetine patients completing sub-chronic treatment as compared to 35 (32%) placebo patients. Table B.I.2 shows the primary reasons for discontinuation during the sub-chronic treatment phase.

The 75 patients who completed the sub-chronic treatment entered the relapse prevention phase. Twenty of these fluoxetine-treated responders were randomized to continue on their current treatment (Flx/Flx) and additional 20 fluoxetine-treated responders were rerandomized to placebo (Flx/Plc). **Among the 20 fluoxetine-treated patients who**

were randomized to continue on their current treatment, only 1 patients received fluoxetine 40 mg/day, the rest of the patients were on 20 mg/day. The sponsor indicated that having only 1 patient at a dose over 20 mg/day does not allow for conclusions to be drawn about higher dose levels. The remaining 35 patients were in the placebo group and continued on treating by placebo. The reasons for discontinuation for the 40 re-randomized patients were summarized in Table B.I.3.

Table B.I.1 Summary of Reasons for Discontinuation (Acute Treatment Phase)

Primary Reason for Discontinuation	Flx (n=109)		Placebo (n=110)		Total (n=219)	
	n	(%)	n	(%)	n	(%)
Complete Period IV	90	(82.6)	68	(61.8)	158	(72.1)
Adverse Event	5	(4.6)	9	(8.2)	14	(6.4)
Lack of Efficacy	5	(4.6)	12	(10.9)	17	(7.8)
Patient decision	3	(2.8)	11	(10.0)	14	(6.4)
Physician Decision	1	(0.9)	0		1	(0.5)
Protocol Requirement	4	(3.7)	3	(2.7)	7	(3.2)
Lost to Follow-up	1	(0.9)	7	(6.4)	8	(3.7)

Table B.I.2 Summary of Reasons for Discontinuation (Sub-chronic Treatment Phase)

Primary Reason for Discontinuation	Flx (n=109)		Placebo (n=110)		Total (n=219)	
	n	(%)	n	(%)	n	(%)
Complete Period V	40	(36.7)	35	(31.8)	75	(34.2)
Adverse Event	11	(10.1)	11	(10.0)	22	(10.0)
Lack of Efficacy	18	(16.5)	23	(20.9)	41	(18.7)
Patient decision	14	(12.8)	16	(14.5)	30	(13.7)
Physician Decision	3	(2.8)	0		3	(1.4)
Protocol Requirement	18	(16.5)	16	(14.5)	34	(15.5)
Lost to Follow-up	5	(4.6)	9	(8.2)	14	(6.4)

Table B.I.3 Summary of Reasons for Discontinuation (Relapse Prevention Phase)

Primary Reason for Discontinuation	Flx/Flx (n=20)		Flx/Plc (n=20)		Total (n=40)	
	n	(%)	n	(%)	n	(%)
Complete Period VI	10	(50.0)	8	(40.0)	18	(45.0)
Adverse Event	1	(5.0)	0	(0.0)	1	(2.5)
Patient Decision	3	(15.0)	0	(0.0)	3	(7.5)
Relapse	6	(30.0)	12	(60.0)	18	(45.0)

Fifteen investigator sites had patients randomized to the acute treatment period. Investigator 6 had no patients randomized to fluoxetine group and only 1 patient randomized to the placebo group. According to the protocol, investigators with no patients randomized into one or more treatment groups were to be pooled in the analysis. So, investigator 6 was pooled with the next lowest enroller, Investigator 17.

For the **acute treatment period**, due to the stratified randomization (at Visit 4), the treatment allocation was balanced within age and gender groups (Table B.I.4). Other baseline characteristics (height and weight) also appear to be comparable. The majority of the patients were Caucasian (82%).

For the **relapse prevention period**, most of the baseline characteristics also appear to be comparable except age and height (Table B.I.5). There were more adolescent patients randomized to Flx/Flx group (60%) as compared to Flx/Plc group (35%). The mean height in Flx/Flx group (160.65 cm) was significantly larger than that in Flx/Plc group (150.53 cm) (p-value=0.019, based on one-way ANOVA with treatment as an independent variable).

With respective to psychiatric history (the age at onset, duration of current episode and previous episodes), the distribution between treatment groups seems similar for both acute treatment and relapse prevention periods, except that the duration of current episode in the Flx/Flx group looks longer than that of the Flx/Plc for the relapse prevention period (93.8 versus 53.2 weeks).

Family history of depression, anxiety disorders and manic depression (Bipolar) also seem comparable between treatment groups. The majority of the patients (more than or equal to 60%) had a family history of depression.

More than or equal to 80% of the patients did not have previous treatment for depressive disorder in both treatment periods.

Table B.I.4 Baseline Patients Characteristics

(Acute Treatment Period)

Variables		Fluoxetine (n=109)	Placebo (n=110)	Total (n=219)
Age	Mean	12.70	12.69	12.70
	Median	12.56	12.50	12.56
	SD	2.46	2 67	2.56
	Min.	8.26	8 01	8.01
	Max.	17 52	17.85	17.85
Age category				
	8 - <13 yrs.	61 (56 0)	61 (55.5)	122 (55.7)
	13 - <18 yrs.	48 (44.0)	49 (44.5)	97 (44.3)
Gender	Female	54 (49.5)	54 (49.1)	108(49.3)
	Male	55 (50.5)	56 (50 9)	111(50 7)
Race	White	96 (88.1)	84 (76.4)	180(82.2)
	Asian	1 (0.9)	0	1 (0.5)
	Black	6 (5.5)	8 (7.3)	14(6.4)
	Hispanic	3 (2.8)	10 (9.1)	13 (5.9)
	Other	3 (2 8)	8 (7.3)	11(5.0)
Weight (kg)				
	N	108	110	218
	Mean	57.06	56.85	56.96
	Median	53.30	55.57	54.43
	SD	19.33	19.96	19.61
	Min.	20.87	26.76	20.87
	Max.	102.97	140.61	140.61
Height (cm)				
	N	108	110	218
	Mean	155.53	153.98	154.75
	Median	156.21	154.94	154.94
	SD	14.55	13.05	13.80
	Min.	124.46	121.92	121.92
	Max.	187.96	182.88	187.96
CDRS-R Total				
	N	109	110	219
	Mean	57.1	55.4	56.2
	Median	55.0	55.0	55 0
	SD	9.9	11.6	10 8
	Min.			
	Max.			

Table B.I.5 Baseline Patients Characteristics
(Relapse Prevention Period) *

Variables		Flx/Flx (n=20)	Flx/Plc (n=20)	Total (n=40)
Age	Mean	13.45	11.65	12.55
	Median	13.70	10.90	12.25
	SD	2.38	2.48	2 57
	Min.	8 91	8.26	8.26
	Max.	17.52	16.27	17.52
Age category				
	8 - <13 yrs	8 (40.0)	13 (65.0)	21 (52 5)
	13 - <18 yrs.	12 (60.0)	7 (35.0)	19 (47.5)
Gender	Female	9 (45.0)	11 (55.0)	20(50.0)
	Male	11 (55.0)	9 (45.0)	20(50.0)
Race	White	17(85.0)	20 (100)	37(92.5)
	Asian	1 (5.0)	0	1 (2.5)
	Black	2 (10.0)	0	2(5.0)
Weight (kg)				
	N	20	20	40
	Mean	61.85	58.35	60.10
	Median	60.78	54.88	57.15
	SD	18.42	21 48	19.83
	Min.	34.02	25.85	25.85
	Max.	102.97	96 16	102.97
Height (cm)				
	N	20	19	39
	Mean	160.65	150.53	155.72
	Median	162.56	147.32	157.48
	SD	11.42	14.19	13.67
	Min.	142.24	124.46	124.46
	Max.	182.88	175.26	182.88
CDRS-R Total				
	N	20	20	40
	Mean	21.9	24.0	22.9
	Median	22.0	25.5	22.5
	SD	3.4	3.7	3 7
	Min.			
	Max.			

Note : * Baseline for age, gender and race was defined as Visit 1; baseline for height was defined as Visit 15.

In **relapse prevention period**, more than 80% of the patients had concomitant medications. In the **acute treatment period**, 82% of the fluoxetine-treated patients took concomitant medications, while only 66% placebo-treated patients took concomitant

medication. In both periods, the most commonly used concomitant medication was the over-the-counter, non-steroidal anti-inflammatory medications, paracetamol and ibuprofen.

Baseline scores for CDRS-R, CGI-Severity, MADRS and MAMA were evaluated at acute treatment period baseline (Visit 4) and relapse prevention baseline (Visit 15). Most of these scores were similar between treatment groups with the following exceptions:

- The CGI-severity **acute-treatment baseline scores** for fluoxetine-treated group were significantly higher than those of the placebo-treated group (mean score of 4.5 vs. 4.4 for fluoxetine-treated and placebo-treated groups, respectively);

- The HAMA **relapse prevention baseline scores** for the flx/plc group were significantly higher than those of the flx/flx group (means score of 1.7 vs. 3.5 for flx/flx and flx/plc groups, respectively).

In either case, the sponsor indicated that the difference was not considered to be clinically meaningful.

In the primary efficacy analysis for the acute treatment period, the sponsor found a greater percentage of response (30% or greater decrease in CDRS-R total score) achieved in the fluoxetine-treated group (65%) as compared with the placebo group (54%). However, the difference was not statistically significant (p-value=0.093) for the 9-week treatment period (Table B.I.6).

The sponsor found statistically significantly greater percentage of patients achieving response than the placebo arm for percent change criteria of 10% (p=0.03), 20% (p=0.002), 40% (p=0.002), 50% (p=0.007) and 60% (p=0.013).

Fluoxetine treated patients were also observed to have higher percentage of remission (41%) as compared with placebo-treated patients (20%) (p<0.01), higher percentage of CGI-improvement response (score of 1 or 2) (52%) compared with placebo (37%) (p=0.028) and higher recovery rate (39%) compared with placebo (20%) (p<0.01) (see Table B.I.8).

In addition, the sponsor also found fluoxetine-treated patients experienced a mean reduction in CDRS-R total score of 22 points as compared to 15 points reduction of placebo group (p<0.001 based on one way ANOVA with treatment as the only factor).

The sponsor performed subgroup analyses on the primary endpoint to examine the consistency of treatment effect across various subgroups : age (8 to <13, 13 to <18 years old), gender, and family history of depression (positive, negative). Based on the Breslow-Day test for homogeneity of odds ratios, they found that there were no statistically significant differences between treatment groups for age, gender or family history of

depression. Fluoxetine-treated group shows higher percentage of response rate consistently across various subgroups (Table B.I.7).

Table B.I.6 CDRS-R Total Score
Number of patients Meeting Criteria for Response
(Percent reduction from baseline)
Acute Treatment Period

% reduction	N	No N	No %	Yes N	Yes %	p-value
30%						
Flx 20 mg	109	38	34.9	71	65.1	2-sided Fisher's exact : 0.093
Placebo	101	47	46.5	54	53.5	Pearson Chi-square : 0.085 Cochran-Mantel-Haenszel : 0.077 ♣

Note : ♣ : Stratification factor included. This reviewer obtained p=0.093 by adjusting for age group (8-<13, >=13 years old) and gender.

Table B.I.7 CDRS-R Total Score by Subgroup
Number of patients Meeting Criteria for Response
(30% reduction from baseline) (Acute Treatment Period)

Subgroup	Therapy	N	Response N	Response %	p-value (2-sided Fisher's exact)
Age					
8 to <13	Flx 20 mg	61	42	69	.128
	Placebo	55	30	55	
13 to ≤18	Flx 20 mg	48	29	60	.533
	Placebo	46	24	52	
Gender					
Female	Flx 20 mg	54	39	72	.397
	Placebo	48	30	63	
Male	Flx 20 mg	55	32	58	.248
	Placebo	53	24	45	
Family history Of depression					
Yes	Flx 20 mg	61	42	69	.339
	Placebo	57	34	60	
No	Flx 20 mg	41	26	63	.348

	Placebo	34	17 50	

Table B.I.8 Categorical Efficacy Analyses – All Randomized Patients

Measure	Acute Treatment (Study Periods III-IV)		
	Flx 20 (%)	Placebo (%)	p-value[a]
CDRS-R total response[b]	65	54	0.093
CDRS-R Total Remission[c]	41	20	<0.01
CGI-improvement response[d]	52	37	0.028
Recovery Rate[e]	39	20	<0.01

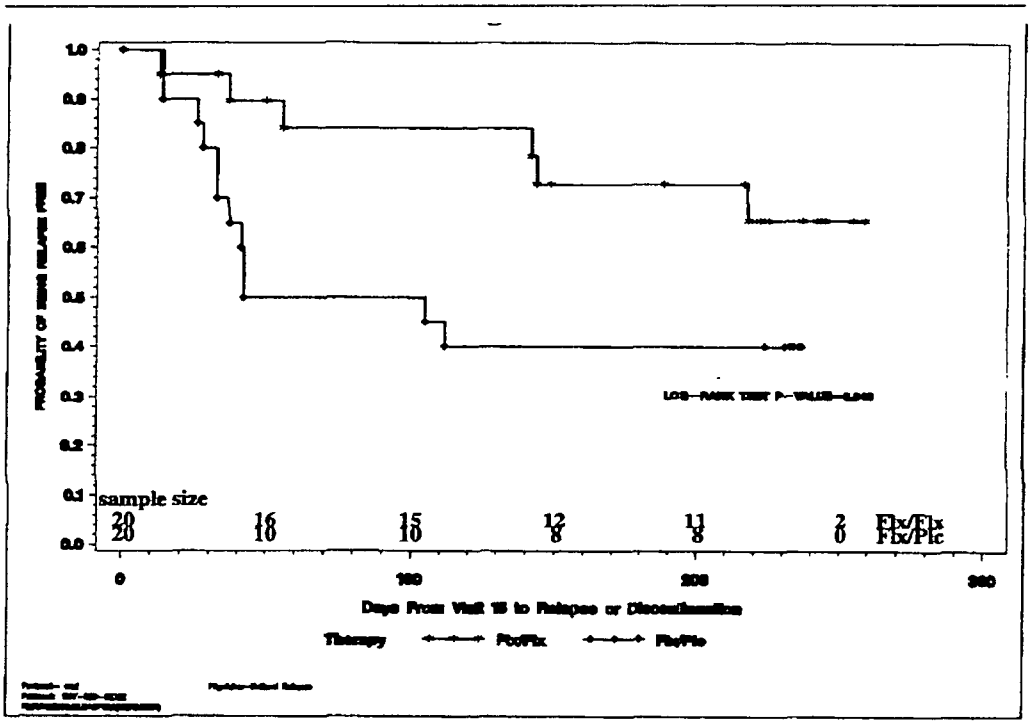
Note : a- Fisher's exact test.
b- ≥30% reduction from baseline.
c- endpoint CDRS-R score ≤28.
d- CGI-improvement score of 1 (very much improved) or 2 (much improved)
e- Recovery – CDRS-R total score ≤28 combined with the CGI-improvement response.

In the time to relapse analysis (protocol-defined or clinician determined relapse), the sponsor found Flx/Flx showed statistically significantly longer time to relapse as compared with Flx/Plc (p-value=0.046). Figure B.I.1 shows the Kaplan-Meier estimates for the time to relapse (primary analysis) during the relapse prevention phase. The estimated percentage of patients meeting the relapse criteria was 34% (95% C.I.=[12%,57%]) for Flx/Flx patients and 60% (95% C.I.=[39%,82%]) for Flx/Plc patients.

As a secondary analysis, the time to **protocol-defined relapse** was analyzed similar to the primary analysis. The more favorable results was found for the Flx/Flx group (p-value=0.032). The estimated percentage of patients met criteria for protocol-defined relapse was 21 % (95% C.I.=[0%,43%]) for Flx/Flx patients and 47% (95% C.I.=[25%,69%]) for Flx/Plc patients.

The sponsor also found that the difference in mean change from baseline (Visit 15) to endpoint (last of Visits 16 through 26) of the CDRS-R total scores for patients entering the relapse prevention phase was not statistically significant between Flx/Flx and Flx/Plc groups (p=0.139).

Figure B.I.1 Kaplan-Meier Survival Estimates for Time to Relapse (protocol or clinician defined) during the Relapse Prevention Phase



III.7 Reviewer's Evaluation and Comments

Similar to study X065, due to the concern of the validity of the sponsor's 30% reduction on CDRS-R total score endpoint, other endpoints will be evaluated as well. This reviewer performed the sponsor's primary analysis and obtained the same result which concluded that fluoxetine group had a favorable but non-statistically significant treatment effect based on percentage of patients who had ≥30% reduction on CDRS-R total scores (p=0.093) from baseline to visits 6 to 10 (acute treatment phase). This reviewer added additional stratified factor (age group or gender) to the Cochran-Mantel-Haenszel (CMH) test, the result remained non-significant (p=0.088 for CMH stratified by age group and p=0.091 for CMH stratified by gender). Though the sponsor did not win on the protocol specified primary endpoint, the sponsor appeared to achieve nominal significance on the other secondary endpoints (CDRS-R Total Remission, CGI-improvement response,

Recovery Rate, mean change from baseline of the CDRS-R total score) as well as the response rate if the other cut-off points were used instead (e.g. 20%, 40%).

At phase V, 29 fluoxetine-treated patients who were non-responders were randomized to either obtain titrated dosing (40 mg-60 mg) or stay at the original dose (20 mg). At this phase, fluoxetine-treated patients and placebo patients were not treated equally, therefore the treatment comparison between fluoxetine-treated group and placebo during the sub-chronic period may not be appropriate. After the titration, only approximately one fifth of the acute treatment period ITT population (20 Flx/Flx and 20 Flx/Plc out of a total of 219 ITT patients) were eligible to be included in the relapse prevention phase. Since there was no decision rule proposed in the protocol for the relapse prevention analysis and the sponsor did not win on the primary efficacy endpoint of the acute treatment period, the performance of additional (i.e. relapse prevention) analysis may not be appropriate. Therefore, the inclusion of the relapse prevention result as a claim is not recommended.

However, this reviewer had confirmed sponsor's time to relapse analysis result in favor of Flx/Flx group (p=0.046). A summary of analyses based on the protocol defined, clinician determined relapse or either method define relapse was presented in Table B.II.1. Note that the relapse defined by protocol or clinician was the primary endpoint indicated in the protocol for the relapse prevention phase.

Since the relapse prevention result was marginal (p=0.046), this reviewer performed a sensitivity analysis to confirm the robustness of the result by setting each of those early withdrawn patients (see Table B.I.3) as having a relapse event (a worst case). With four early withdrawn patients in Flx/Flx group (note: no early withdrawn patients were in Flx/Plc group) reset to have relapse events, the result became non-significant (p=0.267). This demonstrates that the result is not robust since with small change of the event classification based on the data, the result was changed completely.

Table B.II.1 Summary of Analysis Result based on Different Relapse Definition

Relapse defined by	Therapy	N	Relapse N (%)	p-value (Log-rank Test)
Protocol or clinician (Primary)	Flx/Flx	20	6 (30)	0.046
	Flx/Plc	20	12 (60)	
Protocol	Flx /Flx	20	3 (15)	0.032
	Flx/Plc	20	9 (45)	
Clinician	Flx/Flx	20	3 (15)	0.673
	Flx/Plc	20	3 (15)	

IV. Study B1Y-MC-HCJW (OCD)

This study was conducted from March 17, 1999 to February 1, 2000. Children between ages 7 to <13 and adolescent between ages 13 to <18, who had a primary psychiatric diagnosis of Obsessive-Compulsive disorder (OCD : defined by the DSM-IV criteria) were included.

Other inclusion criteria included : OC symptoms of at least moderate severity at Visits 1 and 2, as defined by a rating of moderate or greater (≥ 4) on the CGI-Severity Scale and by a score ≥ 16 on the CY-BOCS (one of the symptoms must have been present for at least 6 months prior to study entry); Baseline score ≥ 7 on the NIMH Global OCD scale; Child Depression Rating Scale ≤ 40 at study entry; without abnormal laboratory or ECG findings; patients who can communicate comprehensibly with study personnel; patients and the principal care givers were judged to be reliable and agree to follow the study procedure.

IV.1 Study Design

This was a multicenter, randomized, double-blind, parallel, placebo-controlled study comparing the efficacy and safety of fluoxetine and placebo in the treatment of children and adolescent patients with OCD.

The study consists two study periods with a total of 10 visits :

- Period I was a 1-week (3-14 days) screening phase for evaluation of study eligibility.
- Period II was a 13-week, double-blind acute treatment period during which patients were randomized (at Visit 2) to fluoxetine or placebo (in a ratio of 2:1).

Patients who were randomized to fluoxetine receive fluoxetine 10 mg daily for 2 weeks, then 20 mg daily for 2 weeks. After these two weeks, the dose may be increased to 40 mg or 60 mg daily based on response and tolerability. If the dose was not tolerated, the 40 mg or 60 mg daily dose may be reduced to 20 and 40 mg daily, respectively. Once the dose was reduced, the patients must remain on that dose for the rest of the study. Patients who can not tolerate 20 mg will be discontinued.

IV.2 Objectives

The primary objects of the study was

- To test the hypothesis that fluoxetine 20 to 60 mg daily is more effective than placebo in the acute treatment of children and adolescents with OCD during the 13 weeks of double-blind treatment period, based on the Children's Yale Brown Obsessive Compulsive (CY-BOCS) total score.

IV.3 Efficacy Endpoint

The primary efficacy measure was the change from baseline to endpoints in total CY-BOCS scores. The CY-BOCS is a clinician-rated instrument designed to measure the presence and severity of OCD symptoms in children (Goodman, 1989b) which was modeled after the Yale-Brown Obsessive Compulsive Scale from adult. It contains 10 items on a 0 to 4-point scale. Five of the items are for obsessive symptoms and the other 5 items are for compulsive symptoms. Two additional questions were collected as part of the assessment; but these items were not scored.

A categorical definition of responder is a patient who has greater than or equal to a 40% reduction in the CY-BOCS total score from baseline to endpoint.

The secondary efficacy measures include

- Clinical Global Impression of Severity (CGI-severity) : a clinician-rated instrument on a 7-pooint scale (1: very much improved and 7 : very much worse);
- Clinical Global Impression of Improvement (CGI-Improvement) : a clinician-rated scale to compare therapeutic effect of treatment versus the conditions at baseline. It is on a 7-point scale (1: very much improved; 4 : no change and 7 : very much worse). This scale was also rated by the patients' parent/guardian;
- Multidimensional Anxiety Scale for Children (MASC) : a self-report scale consisting of 39 items across 4 major factors : physical, social anxiety, harm avoidance and separation anxiety;
- Patient Global Impression (PGI) : a patient-rated perception of changes from start of therapy. It is a 7-point scale (1: very much improved and 7 : very much worse);
- National Institute of Mental Health (NIMH) Global OCD Scale : scale consisting of a rating from 1 to 15. It is used to reflect the severity of a patient's OCD with higher score indicating severe condition;
- Children's Depression Rating Scale-Revised (CDRS-R) : modeled after the Hamilton Depression Rating Scale (HAMD) for adults. The scale consists of 17 items scored on a 1 to 5 or 1 to 7 scale (ranged from 17 to 113, with 1 indicates normal).

Only CY-BOCS, NIMH Global OCD Scale and CGI-Severity were conducted at every visit and upon discontinuation.

IV.4 Number of Subjects and Analysis Plan

A sample size of 90 patients was calculated to achieve 80% power and to detect 4.8-unit difference in mean Y-BOCS total score between fluoxetine and the placebo group, assuming a standard deviation of 7.6. Assuming 10% drop-out rate, a sample size of 100 patients was planned for this study.

All efficacy and safety analyses were conducted based on the intent-to-treat principle. An intent-to-treat analysis is an analysis by the groups to which patients were assigned by random allocation, regardless what patients took later or whether patients followed the protocol.

All tests were performed at a 2-sided 0.05 significance level. **Investigators with fewer than 2 randomized patients per treatment group were pooled for analysis.**

Patients who had at least one post-randomization data were included in the analysis. For change from baseline measures, patients who had a baseline scores (latest of Visit 1 or Visit 2 score) and at least one post-baseline score were included. Total scores were derived from the individual items. If any individual item was missing, then the total score was treated as missing.

The last-observation-carried forward (LOCF) approach was used for the primary efficacy analysis. The primary efficacy analysis was based on the analysis of variance (ANOVA) model with treatment, investigator and treatment-by-investigator effect as independent factors. If the interaction effect was not significant at 0.10 level, the effect was dropped. The statistical inference was based on the F-test on treatment effect based on ANOVA.

A repeated measure analysis of variance was conducted on the CY-BOCS total score to evaluate the treatment effect over time. The dependent variables was the baseline and all the post-baseline CY-BOCS total scores. The model included fixed class effect terms for treatment, investigator, treatment by investigator interaction, visit, and an interaction term of treatment by visit. If the treatment by investigator interaction was not significant at 0.10 level, the interaction term will be excluded from the model. A single degree-of-freedom contrast of comparing the difference between baseline to Visit 10 scores across treatment groups was evaluated. Inference from the repeated measure analysis was based on the restricted maximum likelihood method and from its approximate F-test. Four different within patient covariance structures were considered: unstructured, heterogeneous toeplitz, heterogeneous autoregressive of order 1, and heterogeneous compound symmetry. The covariance structure will be chosen based on the information criteria.

Other analyses of continuous secondary parameters were similar to the primary analysis. For proportion of responders, an exact Mantel-Haenszel test with investigator as the stratification variable will be performed. In the study report, the sponsor provided additional analyses based on the proportional odds model to analyze three endpoint variables (CGI-improvement, parent-rated CGI-improvement and PGI-improvement). The sponsor claimed that the model can preserve the ordinal categorical nature of the

data. Only treatment was included as an independent variable in the proportional odds model.

The primary efficacy measures, CY-BOCS total score was also evaluated by several subgroups such as gender and age (7 to <13 and ≥ 13 to <18 years). Other subgroups will be identified based on the results of baseline variable comparison between treatment groups. The baseline variable comparison will be performed based on ANOVA model with treatment, baseline variables and treatment by baseline variable interaction. Any presence of significant treatment by baseline variable interaction (at 0.05 level) from the ANOVA model will warrant a subgroup analysis.

IV.5 Sponsor's Results

A total of 103 patients were randomized at Visit 2 : 71 to the fluoxetine group and 32 to the placebo. Sixty-nine patients (67%) completed the study with 49 (69%) in the fluoxetine-treated group, compared with 20 (63%) in the placebo-treated group. The primary reason for discontinuation was lack of efficacy (14% and 25% for fluoxetine and placebo groups, respectively) (Table C.I.1).

Table C.I.2 shows the distribution of patient demographic and baseline characteristics by treatment groups. There were similar percentage of girls (52%) and boys (48%). The gender distribution appeared to be compatible between treatment groups. The majority of the patients were Caucasian (86%). Patients had mean age of 11.4 years (ranged from 7 to 17.9). There were three time as many children (73%) as adolescents (27%) in the study. But the distribution of the children and adolescents appeared to be similar across treatment groups. The baseline severity based on CY-BOCS total scores seems to be similar between treatment groups. The distributions of weight and height also appear to be compatible between treatment groups.

Table C.I.1 Summary of Reasons for Discontinuation
All Randomized Patients

Primary Reason for Discontinuation	Flx (n=71) n (%)	Placebo (n=32) n (%)	Total (n=103) n (%)

Protocol complete	49	(69.0)	20	(62.5)	69	(67.0)
Adverse Event	6	(8.5)	2	(6.3)	8	(7.8)
Lack of Efficacy	10	(14.1)	8	(25.0)	18	(17.5)
Patient decision	3	(4.2)	0		3	(2.9)
Physician Decision	1	(1.4)	1	(3.1)	2	(1.9)
Protocol Requirement	1	(1.4)	0		1	(1.0)
Lost to Follow-up	7	(1.4)			2	(1.9)

There were a total of 21 investigators who randomized patients in this study. As specified in the protocol, any center with fewer than 2 randomized patients per treatment group were pooled for statistical analysis purpose. In the study, investigators 1, 4, 7, 8, 9, 12, 13, 17, 19, 20, 21 and 22 were pooled which resulted in 10 new investigator sites.

The sponsor compared baseline patient menstruation status and patient habits (smoking, use of smokeless tobacco, alcohol use and caffeine use). Twenty seven percent of fluoxetine-treated group and 35.3% of placebo group had experienced onset of menses. The majority of the patients did not smoke or did not have alcohol consumption (above 99%), but over half of the patients consumed caffeine (59.2% and 65.5% for fluoxetine and placebo groups, respectively).

The sponsor also summarized the patient psychiatric histories at baseline. The mean duration of illness prior the study entry was approximately 5 years (ranged from 5.8 months to 14.7 years). Only very few patients presented with secondary comorbid psychiatric disorders at baseline :

- 5.6% and 3.1% of major depressive disorder in fluoxetine and placebo treated groups, respectively;
- 4.2% of depressive disorder in fluoxetine-treated groups (but not in the placebo group);
- 2.8% attention deficit hyperactive disorder in fluoxetine treated group; and
- 3.1% of dysthymic disorder in placebo treated group (but not in the fluoxetine-treated group).

In comparing the family history of psychiatric condition, there were substantial number of patients reported family history of depression and/or OCD. Thirty eight percent of fluoxetine treated patients and 56% of placebo patients had a family history of depression. Thirty four percent of fluoxetine group and 28% of placebo group had a family history of OCD.

Only 35% of the patients reported one or more historical diagnoses. With regard to previous treatment, the majority of patients (72%) **did not** received treatments within the past year. The most often used treatment were paroxetine (8%), methylphenidate (7%), fluoxetine (6%) and sertraline (5%).

Table C.I.3 shows the analysis results for the primary efficacy variables : the change from baseline of CY-BOCS scores. From the primary analysis, fluoxetine-treated group shows significantly more favorable result as compared with placebo-treated group (p=0.026). Based on the repeated measure model, the treatment difference was also significant (p=0.034) in favor of fluoxetine. Note that the repeated measure model analysis was not based on change-from-baseline score as the dependent variable but based on all the scores taken over time and the reference was based on the contrast of comparing the difference between baseline to Visit 10 scores across treatment groups. In all these models, none of

Table C.I.2 Baseline Patients Characteristics
All Randomized Patients

Variables	Fluoxetine (n=71)	Placebo (n=32)	Total (n=103)

APPEARS THIS WAY
ON ORIGINAL

Age	Mean	11.42	11.41	11.42
	Median	11.07	11.60	11.27
	SD	2.95	2.79	2.89
	Min.	7.03	7.01	7.01
	Max.	17.72	17.93	17.93
Age category n(%)				
	7 - <13 yrs	51(71.8)	24(75)	75(72.8)
	13 - <18 yrs	20(28.2)	8(25)	28(27.2)
Gender n(%)				
	Female	37 (52.1)	17 (53.1)	54(52.4)
	Male	34 (47.9)	15 (46.9)	49(47.6)
Race	White	62 (87.3)	27 (84.4)	89(86 4)
	Asian	0	1 (3.1)	1 (1 0)
	Black	2 (2.8)	0	2 (1.9)
	Hispanic	4 (5.6)	3 (9 4)	7 (6.8)
	Other	2 (4.2)	1 (3.1)	4 (3 9)
Weight (kg)				
	Mean	46.19	41.55	44.75
	Median	41.28	40.82	41.28
	SD	20.87	13 64	18.97
	Min.	20 87	19.96	19.96
	Max.	102.06	77.11	102.06
Height (cm)				
	N	69	32	101
	Mean	146.67	144.41	145.95
	Median	144.78	144.78	144.78
	SD	16.55	14.31	15.84
	Min	116.84	116.84	116.84
	Max.	185.42	172.21	185.42
	Unspecified	2	0	2
CY-BOCS Total Score				
	N	71	32	103
	Mean	24.5	26.3	25.0
	Median	24.0	25.5	25.0
	SD	5.1	4.6	5.0
	Min.			
	Max.			

the investigator by treatment interaction was significant. Therefore, the investigator by treatment interaction term was taken out of the models.

With respect to CY-BOCS sub-scale (Table C.I.3), there was statistical significant treatment difference in the CY-BOCS compulsions score in favor of fluoxetine (p=0.027), but treatment difference was not significant in obsessions sub-scale (p=0.122).

The sponsor performed sub-group analysis and found that fluoxetine-treated group had consistently more reduction in CY-BOCS change from baseline scores across various subgroups as compared with placebo (Table C.I.4).

The sponsor also found significant fluoxetine treatment effect based on the secondary endpoints : endpoint analysis of clinician and parent rated CGI-improvement score, PGI-improvement score, mean change from baseline in CGI-severity score, NIMH global OCD rating. However, the analysis of mean change from baseline of CDRS-R (depression) total score and MASC (anxiety) scales were not significant. The sponsor indicated that the non-significance result was not surprising since the level of depression or anxiety symptom in these patients were in the low to normal range.

Table C.I.3 CY-BOCS Scores-Change from Baseline to Endpoint of CY-BOCS Total Scores and sub-score (All Randomized Patients)

Change from Baseline	Fluoxetine (n=71)	Placebo (n=32)	P-value @ Based on the Primary Analysis	P-value ♣ Based on Repeated measure model
Total Score				
Mean	-9.5	-5.2	0.026	0.034
Median	-9.0	-4.5		
SD	9.2	7 4		
Obsession Sub-score				
Mean	-4.7	-3.0	0.102	0.122
Median	-4.0	-3.0		
SD	4.8	4.3		
Compulsions Score				
Mean	-4.8	-2.2	0.015	0.027
Median	-4.0	-1.0		
SD	5.0	4.5		

Note : @: Treatment effect based on ANOVA model:
Change from baseline in CY_BOCS score=Investigator + treatment
♣ : Test based on contrast of comparing the difference between baseline to Visit 10 scores across treatment groups using the Repeated Measure Model assuming unstructured variance-covariance structure :
CY-BOCS score=Investigator+Treatment+Visit+Treatment x Visit;
Note : CY-BOCS score was based on the data for each visit, not the change from baseline Scores

Table C.I.4 CY-BOCS Total Score by Subgroup
Change from Baseline to Endpoint

Subgroup	Therapy	N	Change from baseline Mean	SD	p-value (model=treatment)
Age					
7 to <13	Fluoxetine	51	-10	9.5	.056
	Placebo	24	-5.7	7.3	
13 to ≤18	Fluoxetine	20	-8.2	8.4	.202
	Placebo	8	-3.6	8.0	
Gender					
Female	Fluoxetine	37	-10.7	9.7	.083
	Placebo	17	-5.9	7.8	
Male	Fluoxetine	34	-8.1	8.5	.140
	Placebo	15	-4.3	7.1	

IV.6 Reviewer's Evaluation and Comments

This reviewer performed the sponsor's primary analysis and obtained the same result which demonstrated that the fluoxetine-treated group had statistically significant reduction in CY-BOCS total scores at endpoint (p=0.026) as compared with placebo group. The significant treatment effect was unchanged when a rank-transformed ANOVA (with treatment and investigator as two factors) was performed (p=0.035). An observed case only analysis at Visit 10 based on ANOVA model with treatment and investigator in the model also showed the significant result in favor of fluoxetine (p=0.04).

In addition, the significant result remains unchanged when gender or age group (<13 and ≥13 years old) was added to the model, individually (p=0.023 when gender was added and p=0.024 when age group was added).

To evaluate whether the treatment effect might be confounded with the CDRS-R (Children's Depression Rating Scale-Revised) total score or MASC (Multidimensional Anxiety Scale for Children) total score, ANCOVA models were fitted including treatment, investigator and either baseline CDRS or MASC total score in the model. The treatment effect remained significant (p=0.023 for both model), but the baseline CDRS total score term or MASC total score term in the model was not significant. This analysis showed that the baseline CDRS-R score and MASC score did not associated with the treatment effect.

V. SUMMARY AND CONCLUSION

V. 1 Pediatric Depression

In summary, for Study X065, based on the original primary endpoint (recovery defined as below 28 on the CDRS-R and a CGI of 1 or 2) from the Dr. Emslie's protocol, the result was not significant (p=0.339). Based on the CGI improvement score (presented as the primary result in Dr. Emslie's '97 paper), although the result was significant (p=0.040), the endpoint had never been designated as the primary endpoint in the protocol.

Nevertheless, in the study report, sponsor obtained a significant result in favor of fluoxetine (p=0.013) based a post-study designated endpoint ≥30% reduction in CDRS-R total scores from baseline). The sponsor achieved nominal significance on the mean change in CDRS-R total scores from baseline (p=0.002).

The sponsor provided a rationale of selecting the 30% reduction on CDRS-R total score from baseline as the cut-off point (described in "section II.2. Objectives" for study X065 of this review) for the primary efficacy endpoint. The clinical significance of this endpoint should be a clinical judgement.

In conclusion, the sponsor did not win on the protocol specified endpoint (i.e. recovery rate), although the sponsor achieved nominal significance based on the post hoc endpoint (≥30% reduction on CDRS-R total score) as well as the secondary endpoint : mean change from baseline of the CDRS-R total score.

From Study HCJE, the sponsor's primary endpoint (30% reduction in CDRS-R total scores from baseline) was pre-specified. But the result for the acute treatment period was not significant (p=0.093). So, based on this primary endpoint, there was no evidence of the treatment effect. The analysis of the mean change from baseline CDRS-R total scores achieved nominal significance (p<0.001).

Overall speaking, the sponsor did not win on these two pediatric depression studies based on the protocol specified primary endpoint. The evidence for efficacy based on the pre-specified endpoint is not convincing. Study X065 achieved nominal significance based on the post hoc endpoint (≥30% reduction on CDRS-R total score) which was not an acceptable endpoint by the DNDP (according to the April 12, 1999 letter to the sponsor). Both studies achieved nominal significance based on the secondary endpoint : mean change from baseline CDRS-R total score.

V.2 Pediatric Depression -- Relapse Prevention

In Study HCJE, after the fluoxetine non-responders in the acute treatment period were re-randomized to receive titrated doses or staying in 20 mg, the fluoxetine responders from this period were then enrolled into a relapse prevention phase. In the analysis of time to relapse, the sponsor concluded a marginal significant relapse prevention effect of fluoxetine (p=0.046) based on only 1/5 of the original acute treatment period ITT

population. This reviewer found that the interpretation of the relapse prevention effect may not be appropriate given that the primary efficacy analysis result in acute treatment period was not significant and there weren't any pre-specified decision rule implemented in this study. In addition, this reviewer found that the result was not robust based on a sensitivity analysis. Therefore, this reviewer would not recommend the inclusion of the relapse prevention result into the labeling claim.

V.3 Obsessive-Compulsive Disorder

From Study HCJW, this reviewer found that the sponsor's result of significant fluoxetine effect on reduction in CY-BOCS total scores at endpoint (p=0.026) was quite robust. The significant treatment effect was not affected by gender or age subgroups. The results were also not changed based on different statistical analysis methods or adjust for baseline CDRS-R or MASC total scores.

Yuan-Li Shen, Dr. PH
Mathematical Statistician

Concur :

Dr. Jin _____ Dr. Chi _____

CC:
NDA: 18-936/SE5-064
HFD-120/Dr. Katz
HFD-120/Dr. Laughren
HFD-120/Dr. Mosholder
HFD-120/Mr. David
HFD-710/Dr. Chi
HFD-710/Dr. Jin
HFD-710/Dr. Shen

Note : This document was saved in c:\nda2000\prozac_pediatric\prozac_ped_stat.doc

---

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Yuan-li Shen
7/17/01 04:03:44 PM
BIOMETRICS

Kun Jin
7/19/01 04:33:08 PM
BIOMETRICS

George Chi
7/20/01 01:07:18 PM
BIOMETRICS