CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

022526Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA #:	022526 / N0062
Drug Name:	Addyi™ (Flibanserin 100 mg q.h.s.)
Indication(s):	Treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women
Applicant:	Sprout Pharmaceuticals, Inc.
Date(s):	Submission: 02/18/2015 PDUFA: 08/18/2015
Review Priority:	Resubmission class2
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Kate Dwyer, Ph.D.
Concurring Reviewers:	Mahboob Sobhan, Ph.D.
Medical Division:	Division of Bone, Reproductive and Urologic Products, HFD-580
Clinical Team:	Catherine Sewell, MD, Clinical Reviewer Olivia Easley, MD, Clinical Reviewer Christina Chang, MD, Clinical Team leader
Project Manager:	Jennifer Mercier

Keywords: NDA review, ROC analysis

This submission pertains to Sprout Pharmaceutical's response to deficiencies listed in September 27, 2013 complete response letter, as recommended in the formal dispute resolution denial letter of February 7, 2014. This submission did not contain new efficacy information but included safety information with regards to driving study.

The original NDA 22-526 was submitted by Boeringer Ingelheimon on October 27, 2009 and received a Complete Response (CR) Letter on August 27, 2010. Sprout Pharmaceuticals, the current Applicant, acquired all rights to flibanserin from Boeringer Ingelheim in 2012 and resubmitted this NDA on March 29, 2013. A second CR letter was issued by the Division on September 27, 2013. An advisory committee (AC) meeting was held on June 4, 2015 to discuss risk/benefit of flibanserin. This review pertains to post-hoc exploratory analyses of efficacy data.

To support the AC meeting for this resubmisson, I conducted additional exploratory subgroup and responder (using ROC method) analyses based on all three phase 3 trials. Post-hoc analyses to explore whether the treatment difference between flibanserin and placebo varies by severity of baseline SSEs, FSFI desire score, and FSDS-R 13 distress score. No notable differences were identified among any subgroups evaluated.

Responder analysis using the patient global improvement (PGI) anchoring question and the observed sexual events showed modest statistically significant treatment benefit in all three studies. The absolute difference in the percentage of responders between flibanserin and placebo across all three endpoints SSEs, FSFI desire score, and FSDS-R 13 distress score was about 9-15%.

From a statistical perspective, although the efficacy of flibanserin, was modest and statistically significant, the overall treatment benefit of flibanserin should be considered in light of the safety profiles seen across all studies.

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

/s/

KATE L DWYER
06/26/2015

MAHBOOB SOBHAN
06/29/2015

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA #:	022526 / N0039
Drug Name:	Flibanserin 100 mg q.h.s.
Indication(s):	Treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women
Applicant:	Sprout Pharmaceuticals, Inc.
Date(s):	Submitted: 03/29/2013 PDUFA: 09/30/2013
Review Priority:	Resubmission
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Kate Dwyer, Ph.D.
Concurring Reviewers:	Mahboob Sobhan, Ph.D.
Medical Division:	Division of Bone, Reproductive and Urologic Products, HFD-580
Clinical Team:	Daniel Davis, MD, Clinical Reviewer Olivia Easley, MD, Clinical Reviewer Christina Chang, MD, Clinical Team leader
Project Manager:	Charlene Williamson

Keywords: NDA review, analysis of covariance, multiple endpoints, ROC analysis

Table of Contents

LIST OF TABLES

1 EXECUTIVE SUMMARY

The Applicant, Sprout Pharmaceuticals, Inc., is seeking approval of flibanserin 100 mg q.h.s. for the treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women. To support this claim, the safety and efficacy data from one phase 3, double-blind, randomized, placebo controlled study 511.147 was submitted. This review is to determine from a statistical perspective if the submitted information supports this claim.

In study 511.147, hereafter referred to as study 147, eligible patients were randomized into the 24-week, double-blind portion of the trial during which they took study medication once daily in the evening. Randomization was stratified by center in a 1:1 ratio to one of the two treatment groups: flibanserin 100 mg q.h.s. and placebo. At least 450 subjects per treatment arm were required to ensure at least 90% power to detect a mean difference of 1 satisfying sexual events and 0.25 of desire score.

To demonstrate efficacy, the following two co-primary efficacy endpoints were pre-specified in the protocol:

- The change from baseline to Week 24 in the score of the Female Sexual Function Index (FSFI©) desire domain.

- The change from baseline in the number of satisfying sexual events (SSE) as measured by the eDiary standardized to a 28-day period.

There were no statistical issues regarding the design and statistical analyses of the efficacy endpoints. However, the following issues were considered during the review of this resubmission:

1. Impact of higher missing values at flibanserin treated group due to adverse events.
2. Changes in the pre-specified analysis methods due to non-normal data.
3. The "Clinical meaningfulness" of small treatment benefit.

To address the impact of missing value on the efficacy conclusion, sensitivity analyses were conducted using baseline carry forward approach in subjects who had no post-baseline data and using subjects who completed the study at week 24 (observed case). To address the method of analyses, distributional assumptions were checked and the results based on the appropriate analyses are reported in this review. Clinical meaningfulness of the modest change was evaluated by ROC analysis.

Our analyses showed statistically significant treatment benefit of flibanserin on the increase in FSFI desire domain and SSE compared to placebo. The treatment benefit with regards to both co-primary endpoints was similar to what was expected at the design stage. In addition, recall period of 7-day vs. 28-day appear to have no effect on the conclusion of the efficacy results. Similar magnitude of small treatment benefit was noted in the responder analyses of clinical meaningfulness, where only 9%-10% subjects reported a treatment benefit of flibanserin.

Although previously conducted study results were not part of this submission, but the results based on the submitted study appear to be similar. From a statistical perspective, the efficacy of flibanserin has been demonstrated, but the clinical meaningfulness of these results should be considered with respect to clinical utility of such a small treatment benefit over the safety profile of this product.

2 INTRODUCTION

2.1 Overview

The Applicant, Sprout, is seeking approval of flibanserin 100 mg q.h.s. for the treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women.

This is a resubmission to original NDA 22-526, submitted on October 27, 2009 and received a Complete Response Letter (CR Letter) on August 27, 2010. One reason for the CR Letter was the lack of substantial evidence in support of flibanserin for the treatment of HSDD in premenopausal women. In order to address this deficiency, an additional phase 3, double-blind, placebo controlled efficacy study, 511.147, was initiated and completed. This study allowed for enrollment of a broader patient population with concomitant medication use than the two previously submitted studies. Additionally, this study included FSFI desire domain as a co-primary outcome measure to replace the eDiary desire as a co-primary endpoint in the previous studies. In addition, study 147 also included a sub-study for assessment of content and validity of FSFI© 28-day recall period. The efficacy evaluation in this review is mostly based on Study 147. Table 1 presents a brief summary of this study.

Table 1: Summary of Study 147

Study Number (No. of Sites / Country) Dates of Study Conduct	Subject Population	Primary Endpoints	Treatments	ITT(FAS)[1]	Design[2]
147 (66 / U.S.) Oct. 2009 to Feb. 2011	Premenopausal Women aged 18 to 55 years with HSDD.	The change from baseline to week 24 in FSFI desire domain The change from baseline to week 24 in the number of SSE	Flibanserin 100 mg Placebo	542 (532) 545 (536)	R, DB, PG, MC, PC

Source: Reviewer's summary based on study reports.

2.2 Data Sources

The study data, reports and additional information for these studies were submitted electronically. The submitted SAS data sets for all studies were complete and well documented. These items are located in the Electronic Document Room at

\\Cdsesub1\evsprod\NDA022526\0039 with submissions dated 03/29/2013.

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

Study 147 was a prospective, multi-center, 24-week, randomized, double-blind, placebo-controlled, two-arm, parallel-group trial comparing the effects of flibanserin 100 mg q.h.s. to placebo in premenopausal women with HSDD.

After a 28-day (+/-7 day) screening period, eligible patients were randomized into the 24-week, double-blind portion of the trial during which they took study medication once daily in the evening. Randomization was stratified by center in a 1:1 ratio to one of the two treatment groups: flibanserin 100 mg q.h.s. or placebo. The overall study schedule is presented in the table below.

Trial Periods	Screen	Baseline	Treatment							End of Treatment[1]	Post Treatment
Clinic Visit	1	2		4	5		7	8	9		
Phone Visit			3			6					10
Week	-4	0	1	4	8	12	16	20	24		25

1. End of Treatment procedures were performed whenever a patient completes or prematurely discontinues from the trial. Patients who prematurely discontinued study medication continued making study visits per protocol; they were not scheduled for End of Treatment procedures except for laboratory tests.

At each visit, the investigative staff assessed medication compliance by counting the tablets remaining. Patients must have taken 80% to 120% of their prescribed study medication to be considered compliant for the entire trial.

Patients were discontinued for:

- Any concomitant illness that prevents compliance, or
- Failure to take any study medication for more than seven consecutive days. These patients would stop taking study drug but continue to attend all study visits and perform all study assessments,
- Pregnancy: if a patient becomes pregnant, she would immediately discontinue the study medication.

A crossover sub-study was nested in study 147 in order to validate the FSFI© 28-day recall period. All patients remaining in the trial that have not yet completed Visit 8 was centrally assigned to two groups in a 1:1 ratio. Each site received a listing from the sponsor identifying each patient's group assignment. One group completed the standard FSFI© with a 28-day recall period and the other group completed a modified FSFI© with a 7-day recall period. The FSFI© must be administered prior to all other evaluations. Sites notified the sponsor on a continuous basis as each patient completed this visit. Once the target of approximately 240 patients has been

achieved, all subsequent patients would have completed the standard FSFI© with a 28-day recall only.

The primary objective of the study 147 was to demonstrate the efficacy and safety of flibanserin 100 mg q.h.s. over 24 weeks of treatment to produce a clinically meaningful therapeutic response in premenopausal women with HSDD. The objective of the nested crossover study was to validate the FSFI© 28-day recall period.

Primary Efficacy Endpoints: There were two co-primary efficacy endpoints pre-specified in the protocol:

1. The change from baseline to Week 24 in the desire score (Two questions of Female Sexual Function Index (FSFI©) desire domain)

2. The change from baseline in the number of SSE's as measured by the eDiary. The calculation of SSE's was standardized to a 28-day period according to the following formula:

Total monthly events = 28 x (sum of the number of events) / (sum of number of days entered). "Satisfying" means gratifying, fulfilling, satisfactory, and/or successful for the patient. The partner's satisfaction is not the subject of this question.

Female Sexual Function Index©

The FSFI© is a brief, multidimensional, self-administered questionnaire for assessing key dimensions of sexual function in women. The scale consisted of 19 items that assessed sexual function over the past four weeks in six domains: desire, arousal, lubrication, orgasm, satisfaction, and pain. The two items in the desire domain were scored from '1' to '5'. The raw scores of the two items were added together and then multiplied by the domain factor of 0.6. Thus, the score of the desire domain ranged from 1.2 to 6.0. For the entire instrument each of the six domains contributed a maximum of '6' points to the total. Thus, the maximum score of FSFI© was '36'.

Secondary Efficacy Endpoints:

Key secondary endpoint:

- The change from baseline to Week 24 on Question 13 of the Female Sexual Distress Scale-Revised (FSDS-R©).

Female Sexual Distress Scale-Revised©

The FSDS© is a measure of female personal distress associated with sexual dysfunction. Reliability and validity of the FSDS© (12-item version), with a 30-day recall period, have been evaluated in different samples of sexually functional and dysfunctional women. For the FSDS©, results indicated a unidimensional factor structure, a high degree of internal consistency, and test–retest reliability. The FSDS© showed a high degree of discrimination between sexually dysfunctional and functional women in each of its three validation studies (R04-1068).

An additional question (Question 13) has been added to the validated FSDS©. This question was about distress specifically related to sexual desire ("bothered by low sexual desire"). FSDS© plus Question 13 comprises FSDS-R© which makes the FSDS-R© a self-administered 13-item questionnaire. The maximum total score of the FSDS-R© indicating the maximum level of sexual distress was '52.'

Additional secondary endpoints:

- The change from baseline in the SSE count (i.e., without standardization to a 28-day period),
- The change from baseline to Week 24 on the FSFI© total score,
- The change from baseline to Week 24 on the FSDS-R© total score,
- The change from baseline in Score on the PGI of Improvement at Week 24.

PGI of Improvement

The PGI of Improvement is a simple evaluation completed by the patient to assess the patient's overall evaluation of her status. The PGI of Improvement was rated ordinal from one to seven where score of 2 means "much improved", 3 means "minimally improved" and score of ≥4 means "no change". In this trial, the PGI of Improvement score was used to assess the patient's evaluation of overall improvement of her HSDD.

Responder Analyses based on PGI and:

FSFI© desire domain,

SSEs (standardized as well as simple count), and

FSDS-R© Question 13.

For each of the above, a responder was defined as a patient with a change from baseline in the endpoint value that was greater than the response threshold defined by the difference between "minimally improved" (score of 3) and "no change" (score of 4) on the PGI of improvement.

3.1.2 Statistical Methodologies

As specified in the protocol, the first co-primary endpoint of change from baseline in the score of the FSFI© desire items was analyzed using analysis of covariance (ANCOVA) with treatment and pooled center as fixed effects, baseline score and hormonal contraceptive use as covariates.

The second co-primary endpoint of change from baseline in the frequency of satisfying sexual events (SSE) measured by the eDiary was analyzed using a stratified Wilcoxon rank sum test where strata were the pooled centers. The adjustment for baseline SSE was taken into account by using change from baseline in the number of SSE in the test.

To declare success, both co-primary endpoints must be statistically significant as pre-specified in the protocol.

The key secondary endpoint of change from baseline to Week 24 in the FSDS-R© Question 13 score was analyzed using ANCOVA with treatment and center as fixed effects and baseline score as a covariate.

To compare the FSFI-SD 28-day recall assessment to the 7-day recall assessment, the following analyses were performed:

1. An equivalence test approach was used via ANOVA to examine the mean differences between the two recall periods taking treatment, sequence and period effect into account in the model. The mean difference between the 7-day and 28-day recall periods was assessed using a value of 0.6 which is a one unit change on the FSFI-SD (also the smallest unit change on the FSFI desire).
2. Cohen's D was calculated to compare the 28-day to the 7-day recall assessment relative to the standard deviation.
3. Examined the ratio (ratio of what? Need to specify) of the within patient assessment of the two different recall periods. This approach takes advantage of the cross-over design looking at the within patient ratio of the 28-day recall to the 7-day recall assessment and determining that the 95% confidence interval of the ratio within 80-125% (note that if the two assessments are exactly the same, the ratio would be 100%).
4. Analyses looking at the correlation between the 28-day and the 7-day recall assessments were also examined. The intraclass correlation coefficient (ICC) was examined.

Two analyses populations were pre-specified in the protocol: (1) the intent-to-treat (ITT), and (2) the full analysis set (FAS). The ITT population consisted of those patients who received at least one dose of study medication. The FAS population consisted of those patients who were randomized to a treatment group, received at least one dose of study medication, and had at least one on-treatment efficacy assessment. In the analysis using FAS population, missing values were handled with last-observation-carried-forward (LOCF) approach and was considered the primary analysis.

For secondary endpoints, ANCOVA or Wilcoxon rank sum test (as deemed appropriate) were used for continuous endpoints and Cochran-Mantel-Haenszel test was used for the responder analysis.

3.1.3 Patient Disposition, Demographic and Baseline Characteristics

In Study 147, a total of 1,090 subjects were randomized into the study (543 in flibanserin 100mg and 543 in placebo group). Of the 1,090 randomized, 3 did not use the test article. Among 1,087 patients who took at least 1 dose of test article, 854 (78.6%) completed the study. Overall, approximately 21% of the subjects discontinued the study prematurely, the main reason being administrative (9.4%) followed by adverse event (6.7%). More subjects in flibanserin 100 mg q.h.s. group discontinued the study early due to "Adverse Event" compared to subjects in placebo group (9.8% vs. 3.7%, respectively). Details of the subject disposition in Study 147 are summarized in Table 2.

Table 2: Subjects Disposition, Study 147

Category	FLI 100mg qhs		Placebo		Total	
Randomly Assigned	543		547		1,090	
Randomized and Treated (ITT)	542	100%	545	100%	1,087	100%
Full Analysis Set (FAS)	532	98.2%	536	98.3%	1,068	98.3%
Completed the Study	408	75.3%	446	81.8%	854	78.6%
Discontinued Study Drug	134	24.7%	99	18.2%	233	21.4%
Reason for Discontinued						
Adverse Event	53	9.8%	20	3.7%	73	6.7%
Administrative	50	9.2%	52	9.5%	102	9.4%
Lost to Follow-up	31	5.7%	28	5.1%	59	5.4%
Non-Compliance with Protocol	15	2.8%	12	2.2%	27	2.5%
Consent Withdraw	4	0.7%	12	2.2%	16	1.5%
Lack of Efficacy	3	0.6%	3	0.6%	6	0.6%
Others*	28	5.2%	24	4.4%	52	4.8%

(Source: Clinical Study 147 Reports; Table 10.1:1 page 61 & Reviewer's Analyses)

The demographics and baseline characteristics were comparable among the treatment groups (Table 3). The subjects mean age was 36.5 years ranging from 18 to 55 years. The majority of subjects were Caucasian (> 85%). The mean present relationship year was 10.9.

Table 3: Demographic and Baseline Characteristics: Study 147 (ITT Population)

Characteristic Mean (SD)	FLI 100mg qhs N=542	Placebo N=545	Total N=1087
Age (years)	36.5 (8.0)	36.6 (7.8)	36.5 (7.9)
18 – 35	221 (40.8%)	222 (40.7%)	443 (40.3%)
35 – 45	223 (41.1%)	225 (41.3%)	448 (41.2%)
45 – 55	98 (18.1%)	98 (18.0%)	196 (18.0%)
Race n (%)			
Black	64 (12.2%)	66 (11.7%)	130 (12.0%)
Caucasian	463 (86.0%)	466 (85.0%)	929 (85.5%)
Other	18 (1.8%)	10 (1.3%)	28 (1.5%)
BMI	27.3 (6.3)	27.3 (7.0)	27.3 (6.7)
Relationship Duration (yrs)	11.1 (7.5)	10.8 (7.2)	11.0 (7.3)

(Source: Clinical Study 147 Reports; Table 11.2:1 page 64& Reviewer's Analyses)

Baseline values for efficacy endpoints for subjects in the FAS populations are shown in Table 4. The primary endpoints of FSFI desire domain, SSEs and secondary endpoints of FSDS item 13, FSDS-R total, FSFI total score SSEs count were similar across treatment groups at baseline.

Table 4: Baseline Characteristic of Efficacy Measures: Study 147 (FAS)

	FLI 100mg qhs			Placebo		
	N	Mean (SD)	Median (Q1, Q3)	N	Mean (SD)	Median (Q1, Q3)
Co-primary endpoints:						
FSFI desire domain	532	1.88 (0.69)	1.8 (1.2, 2.4)	536	1.86 (0.70)	1.8 (1.2, 2.4)
SSEs (standardized)	528	2.51 (2.49)	2 (1, 2)	532	2.66 (2.93)	2 (1, 4)
Key Secondary Endpoint:						
FSDS-R Item 13	532	3.43 (0.69)	3 (3, 4)	536	3.40 (0.67)	3 (3, 4)
Secondary Endpoints:						
FSDS-R Total score	532	32.8 (9.03)	33 (26, 39)	536	32.5 (8.67)	33 (26, 39)
FSFI total score	532	19.0 (6.03)	19.2 (15.3, 23.5)	536	19.0 (6.10)	19.4 (15.2, 23.4)
SSEs count	528	2.40 (2.41)	2 (1, 4)	532	2.57 (2.90)	2 (1, 3)

(Source: Clinical Study 147 Reports; Table 11.2:3, page 66 & Reviewer's Analyses)
*State why N=528 for SSE

3.1.4 Results and Conclusions

3.1.4.1 Primary and Key Secondary Efficacy Endpoints

Results of our analysis on the two co-primary endpoints of changes in SSE and FSFI desire domain and key secondary endpoint of change in FSDS-R Item 13 are presented in Table 5. The change from baseline in these endpoints were not normally distributed, therefore, results based on non-parametric method was reported. At Week 24, treatment with flibanserin resulted in a statistically significant increase in SSE compared to placebo. The median treatment difference was approximately 0.5 (mean difference of 1.0) in SSE, half of what was expected at the design stage. Similarly, the treatment difference for FSFI desire score was 0.60, little more than the hypothesized difference, albeit statistically significantly superior to placebo. However, the clinical significance of this difference was not well understood and needs clinical decision regarding the efficacy of flibanserin. Analysis results using ANCOVA are showed in (Appendix Table 10).

The Applicant performed "responder" analysis to address the clinical significant of modest treatment difference. We also evaluated the clinical meaningfulness of such "minimal" changes in efficacy in the later section.

Table 5: Changes from Baseline at Final Visit using Wilcoxon Test, Study 147 (FAS, LOCF)

	FLI 100mg qhs		Placebo			
	N	Median (Q1, Q3)	N	Median (Q1, Q3)	Treatment Difference.	P-value
FSFI desire domain	506	1.20 (0.0, 1.8)	525	0.60 (0.0, 1.2)	0.60	< 0.0001
SSEs (standardized)	500	1.04 (0.0, 4.0)	521	0.50 (-1.0, 2.2)	0.54	< 0.0001
FSDS-R Item 13	506	-1 (-2, 0)	525	0 (-1.0, 0.0)	-1.0	< 0.0001

Note: P-value based on Wilcoxon rank sum test.
(Source: Reviewer's Analyses)

Sensitivity analyses using baseline carried forward (LOCF0) and completer's population (OC) are presented in Appendix Table 10 and 11. Although the statistical significances were achieved for the treatment difference between flibanserin and placebo for all three efficacy endpoints ($p<0.05$), but the median treatment difference for FSFI desire domain was zero based on either LOCF0 or OC analysis population.

3.1.4.2 Additional Secondary Efficacy Endpoints

Additional secondary endpoints included the change from baseline in the FSFI total score, raw count of all SSEs without standardizing to a 28-day period and FSDS-R total counts. Results of our analysis showed that the treatment differences between flibanserin 100 mg q.h.s. and placebo at Week 24 were statistically significant for all three secondary endpoints (Table 6).

Table 6: Changes from Baseline at Final Visit using ANCOVA for Secondary Endpoints, Study 147

	FLI 100mg qhs		Placebo			
	N	Change from Baseline	N	Change from Baseline	Treatment Difference (95% CI)	P-value
FSFI Total	506	5.29 (0.33)	525	3.49 (0.32)	1.79 (0.98, 2.61)	<0.0001
SSEs (count)	500	2.29 (0.21)	521	1.36 (0.21)	0.92 (0.40, 1.46)	0.0006
FSDS-R	506	-9.35 (0.59)	525	-6.01 (0.59)	-3.28 (-4.74, -1.81)	<0.0001

(Source: Reviewer's analyses)

3.1.4.3 Responder Analyses

The Applicant reported several "cutpoints", based on ROC type analysis to define a subject as responder or non-responder. For the FSFI desire domain, a subject was defined responder if the cutpoints increased by a score of at least 0.6, 1.2, 1.8 and 2.4 points. For SSE (standardized to a 28-day period), a subject was defined as responder if there was an increase in SSE of at least 1, 2, 3, and 4 SSE. For FSDS-R Question 13, responders were pre-defined as a decrease of 1 or more points (out of a 7-point scale) on FSDS-R question 13. Based on these definitions, the

Applicant reported statistically significant difference of 11.6% to 12.4% in favor of flibanserin compared to placebo, with respect to all these three endpoints.

In order to explore further, whether the modest improvement seen in SSE and desire is "clinically meaningful", we also performed a responder analysis based on a "patient global improvement" anchoring question using ROC method. For this analysis, the subjects in the FAS population, irrespective of treatment assignment, were categorized as "satisfied" vs. "unsatisfied" based on the PGI questionnaire results at final visit. The "satisfied" subjects were defined as those whose PGI response was ≤ 3 (much improved to minimally improved), and "unsatisfied subjects" were defined as those whose PGI response was >3 (no improvement) or missing. After categorizing the subjects as "satisfied" and "unsatisfied", a ROC analysis was conducted by fitting a logistic regression model with satisfied vs. unsatisfied as the response and change from baseline in FSFI desire domain, SSE or FSDS-R Item 13 as the covariate, respectively, to determine the "cutoff" point.

Based on the "cutoff" point from the ROC method (PGI≤ 3 as satisfied subjects), the responders were defined as those subjects for whom the mean change was at least 0.6 for FSFI desire domain, 1.57 for SSE and -1 for FSDF-R Item 13, respectively. Subjects were classified as non-responders otherwise. As presented in Table 7, the results showed 9.9% to 12.6% improvement in favor of flibanserin 100 mg q.h.s. at Week 24 compared to placebo. Using the PGI cutoff points of ≤ 2 (much improved), similar results was seen with regards to both co-primary and key secondary endpoints.

Thus, results of responder analyses also showed statistically significant treatment benefit, with a small difference.

Table 7: Responder Analyses based on the "cut-off" from ROC, Study 147 (FAS Population)

	Cutoff	FLI 100mg qhs n/N (%)	Placebo n/N (%)	Treatment Difference	P-value*
FSFI desire domain					
PGI ≤ 3	0.6	308/532 (57.9%)	257/536 (48.0%)	9.9%	0.0019
PGI ≤ 2	1.2	255/532 (47.9%)	203/536 (37.9%)	10.1%	0.0009
SSEs (standardized)					
PGI ≤ 3	1.57	233/532 (43.8%)	180/536 (33.6%)	10.2%	0.0005
PGI ≤ 2	2.89	166/532 (31.2%)	117/536 (21.8%)	9.4%	0.0006
FSDS-R Item 13					
PGI ≤ 3	-1	328/532 (61.6%)	263/536 (49.1%)	12.6%	<0.0001
PGI ≤ 2	-2	180/532 (33.8%)	133/536 (24.8%)	9.0%	0.0015

*P-value is based on Cochran-Mantel-Haenszel test stratified by pooled center.
(Source: Reviewer's analyses)

3.1.4.4 Recall Period Comparison

Results based on the several statistical tests comparing the recall period (7-day vs. 28-day) are presented in Table 8. These results showed that the 7-day and 28-day recall period were similar.

Table 8: Comparison of the FSFI desire domain: 7-day vs. 28-day recall period

Statistic	FSFI 7-day recall vs. FSFI 28-day recall
N	175
Mean Difference (95% CI)	-0.11 (-0.23, 0.01)
Cohen's D	0.20
Ratio (95% CI)	0.90 (0.80, 1.01)
ICC (95% CI)	0.78 (0.72, 0.83)

(Source: Clinical Study 147 Reports; Table 11.4.1.2:5, page 72)

The above results confirmed the Applicant's contention that the level of desire and ratings were no different whether the subjects used 7-day or 28-day recall period.

3.2 Evaluation of Safety

Evaluation of safety data can be found in the clinical reviewer's report.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

Study 147 was conducted in only female subjects, and therefore, analysis by gender and geographical region was not applicable. Subgroup analysis by race was also not applicable because the majority of the subjects were white (>85%).

Results for the age subgroup (< 35, 35 to 40, and ≥ 40) analysis are shown in Table 10. The treatment difference varied across age subgroups and endpoints. For desire domain, the placebo adjusted treatment effects were not different between age groups, while for SSEs treatment effect was more pronounced only in women aged ≥40. Similar results were also noted for FSDS-R Item 13 the treatment was more effective in women aged ≥35.

Table 9: Subgroup Analyses by Agegroup, Study 147 (FAS, LOCF)

	FLI 100mg qhs		Placebo			
	N	Change from Baseline	N	Change from Baseline	Treatment Difference(95% CI)	P-value
FSFI desire domain						
Age < 35	207	1.15 (0.10)	214	0.88 (0.09)	0.27 (0.03, 0.51)	0.0263
35 ≤ Age < 40	89	1.11 (0.14)	107	0.65 (0.13)	0.47 (0.10, 0.83)	0.0123
Age ≥ 40	210	0.77 (0.09)	204	0.51 (0.09)	0.25 (0.03, 0.49)	0.0298
SSEs (standardized)						
Age < 35	204	2.49 (0.37)	210	1.96 (0.36)	0.52 (-0.40, 1.44)	0.2664
35 ≤ Age < 40	88	2.05 (0.55)	107	1.49 (0.50)	0.56 (-0.85, 2.05)	0.4316
Age ≥ 40	208	2.16 (0.35)	204	0.71 (0.35)	1.45 (0.54, 2.37)	0.0019
FSDS-R Item 13						
Age < 35	207	-1.11 (0.10)	214	-0.92 (0.10)	-0.19 (-0.44, 0.06)	0.1299
35 ≤ Age < 40	89	-1.03 (0.16)	107	-0.67 (0.15)	-0.37 (-0.80, -0.04)	0.0746
Age ≥ 40	210	-0.93 (0.09)	204	-0.55 (0.09)	-0.38 (-0.61, -0.15)	0.0015

ANCOVA model: Change = Treatment + Baseline + Site.
(Source: Reviewer's analyses)

5 SUMMARY AND CONCLUSIONS

5.1 Conclusions and Recommendations

The data submitted from study 147 showed that the treatment difference between flibanserin 100 mg q.h.s. and placebo was statistically significant in both pre-specified co-primary endpoints: FSFI desire domain and sexually satisfying events and key secondary endpoint: FSDS-R item 13. There were no statistical issues in terms of analysis methods.

This conclusion was based on the results of both the Applicant and our analyses. The treatment benefits in both co-primary endpoints were similar to what was planned at the design stage. Small treatment benefit was noted in the responder analyses of clinical meaningfulness, where only 9%-10% subjects reported a treatment benefit due to flibanserin. Furthermore, the desire score based on either 7-day recall or 28-day recall period appear to be similar and have no effect on the conclusion of the efficacy results.

Although previously conducted study results were not part of this submission, but the results based on the submitted study appear to be similar. From a statistical perspective, the efficacy of flibanserin has been demonstrated, but the clinical meaningfulness of these results should be considered with respect to clinical utility of such a small treatment benefit over the safety profile of this product.

6 APPENDIX

During the review of this NDA, we noted that approximately 3% of the subjects in the FAS population were not in the LOCF analyses submitted by the Applicant. Among these, there are more subjects missing due to adverse events in flibanserin treated group (3%) than placebo treated group (0.9%). Because these observations are not missing at random, we conducted sensitivity analysis using baseline carry forward (LOCF0) and observed cases (OC). The results of these sensitivity analyses were presented in Table 10 and Table 11.

Table 10: Changes from Baseline at Final Visit using ANCOVA, Study 147

	FLI 100mg qhs		Placebo		FLI vs. Placebo	
	N	Change from Baseline	N	Change from Baseline	Diff. (C.I.)	P-value
FSFI desire domain						
FAS LOCF	506	0.98 (0.06)	525	0.68 (0.06)	0.30 (0.16, 0.44)	<0.0001
FAS LOCF0	532	0.93 (0.05)	536	0.66 (0.55)	0.27 (0.13, 0.40)	0.0001
OC	343	0.95 (0.07)	380	0.66 (0.07)	0.29 (0.12, 0.46)	0.0011
SSEs (standardized)						
FAS LOCF	500	2.43 (0.22)	521	1.46 (0.22)	0.97 (0.43, 1.52)	0.0005
FAS LOCF0	532	2.28 (0.21)	536	1.42 (0.21)	0.86 (0.33, 1.39)	0.0016
OC	329	2.33 (0.27)	361	1.44 (0.25)	0.89 (0.25, 1.55)	0.0070
FSDS-R Item 13						
FAS LOCF	506	-1.00 (0.06)	525	-0.72 (0.06)	-0.29 (-0.43, -0.14)	0.0001
FAS LOCF0	532	-0.96 (0.06)	536	-0.70 (0.06)	-0.26 (-0.40, -0.12)	0.0004
OC	391	-0.98 (0.07)	426	-0.71 (0.07)	-0.28 (-0.44, -0.11)	0.0013

(Source: Reviewer's analyses)

Table 11: Changes from Baseline at Final Visit using Wilcoxon Test, Study 147

	FLI 100mg qhs		Placebo		FLI vs. Placebo	
	N	Median (Q1, Q3)	N	Median (Q1, Q3)	Median Diff	P-value
FSFI desire domain						
FAS LOCF	506	1.20 (0.0, 1.8)	525	0.60 (0.0, 1.2)	0.60	< 0.0001
FAS LOCF0	532	0.60 (0.0, 1.8)	536	0.60 (0.0, 1.2)	0.00	0.0001
OC	343	0.60 (0.0, 1.8)	380	0.60 (0.0, 1.2)	0.00	0.0009
SSEs (standardized)						
FAS LOCF	500	1.04 (0.0, 4.0)	521	0.50 (-1.0, 2.2)	0.54	< 0.0001
FAS LOCF0	532	1.0 (0.0, 3.6)	536	0.35 (-1.0, 2.1)	0.65	0.0003
OC	329	1.0 (0.0, 3.7)	361	0.45 (-1.0, 2.1)	0.55	0.0020
FSDS-R Item 13						
FAS LOCF	506	-1 (-2, 0)	525	0 (-1, 0)	-1	< 0.0001
FAS LOCF0	532	-1 (-2, 0)	536	0 (-1, 0)	-1	< 0.0001
OC	391	-1 (-2, 0)	426	0 (-1, 0)	-1	< 0.0001

(Source: Reviewer's analyses)

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

/s/

KATE L DWYER
08/30/2013

MAHBOOB SOBHAN
08/30/2013

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number:	022526
Drug Name:	Flibaserin Tablet
Indication:	Hypoactive Sexual Desire Disorder (HSDD) in Premenopausal women
Study number:	SPR-12-05
Applicant:	Sprout pharmaceuticals
Date(s):	Date of Document: 3/28/2013 Consult received date: 4/30/2013 PDUFA date: 9/27/2013 Completion date: 8/20/2013
Review Priority:	P
Biometrics Division:	DB VI
Statistical Reviewer:	Ling Chen, Ph.D., Mathematical Statistician, Special Projects Team, DBVI/OB/OTS
Concurring Reviewers:	Yi Tsong, Ph.D., Acting Division Director, DBVI/OB/OTS
Medical Division:	Controlled Substance Staff
The CSS Team:	Katherine R. Bonson, Ph.D., Pharmacologist, OD/CSS Silvia Calderon, Team Leader, Pharmacologist, OD/CSS Michael Klein, Ph.D., Director, OD/CSS
Project Manager:	Sandra Saltz, OD/CSS

Keywords: Crossover design; Drug abuse potential study; Self-reported endpoint; Multiple endpoints

Table of Contents

List of Tables

List of Figures

1. Executive Summary

SPR12-05 was a single-dose, randomized, double-blind, placebo- and active-controlled 6-way crossover study to evaluate the abuse potential of single doses (100 mg, 200 mg, and 250 mg) of flibanserin compared to placebo and 2 doses (15 mg and 30 mg) of zolpidem in recreational polydrug users.

Of the 104 subjects who completed the Qualification Phase, 36 subjects entered the Treatment Phase of the study. Thirty four of the 36 subjects (94.4%) received all 6 treatments and completed all study activities. The reviewer's analysis was based on the 34 completers.

The primary endpoint was Emax of Drug Liking VAS on a bipolar scale. On the average, the responses to each dose of flibanserin were significantly lower than those to zolpidem 30 mg for Drug Liking VAS. The mean differences between two high doses (200 mg and 250 mg) of flibanserin and zolpidem 30 mg were -7.38 and -7.81, respectively. There was no significant difference between two high doses of flibanserin and zolpidem 15 mg. In the comparison of flibanserin versus placebo, on the average, the responses to flibanserin (100 mg, 200 mg and 250 mg) were significantly larger than those to placebo. The difference between each dose of zolpidem and placebo was statistically significant. However, 11 of 34 subjects (32.4%) had an Emax to zolpidem 15 mg less than or equal to 53, and 9 of 34 subjects (26.5%) had an Emax to zolpidem 30 mg less than or equal to 57 for Drug Liking VAS. It means that at least 25% of subjects did not respond to the positive control in the study. Notice that twenty five percent of appropriately selected subjects could change the results from the primary analysis.

The following problems were identified in the reviewer's secondary analysis:

1. Large placebo responses were observed for many secondary measures in the study. For example, 14 of 34 subjects (41.2%) had an Emax of Good Effects VAS to placebo greater than or equal to 39 on the unipolar scale from 0 to 100.
2. For some abuse potential measures, such as High VAS and ARCI measures, predose responses were collected in each treatment period before dosing for each subject in the study. However, these predose responses were not used in the statistical analysis. Instead, the Sponsor inappropriately used baseline responses (defined as predose responses for each subject in Period 1 on Day 1) in calculating the response variable and the covariate in the statistical model for the analysis.
3. Large predose responses were observed in the study. For example, for High VAS on the unipolar scale from 0 to 100, 32 of 34 subjects (94.1%) had predose responses between 46 and 51 in the first treatment period. Because these predose responses are so close to score 50, it appears that subjects may be confused by unipolar and bipolar scales. Similarly, for ARCI MBG (a scale that ranges from 0 to 16, based on 16 specific yes/no questions), in the first treatment period, 12 of 34 subjects (35%) had predose responses between 7 and 14. Among these subjects, 6 subjects had a score 10 or above.
4. Overall Drug Liking VAS and Take Drug Again VAS are on a bipolar scale from 0 to 100. The neutral score is 50. In this study, for Overall Drug Liking VAS 25 of 34 subjects (73.5%) and 24 of 34 subjects (70.6%) had an Emax less than 55 for zolpidem 15 mg and 30 mg, respectively, and for Take Drug Again VAS 22 of 34 subjects (64.7%) and 21 of 34 subjects (61.8%) had an Emax less than 55 for zolpidem 15 mg and 30 mg, respectively.

In this reviewer's opinion the study subjects were not well selected and well trained. The problems listed above make the results from the study not interpretable. Therefore, the study is inconclusive.

2. Review Report on Study SPR-12-05

2.1 Overview

2.1.1 Objectives of the study

Primary objectives

The primary objectives of this study were 1) to evaluate the abuse potential of flibanserin compared to placebo, 2) to evaluate the abuse potential of flibanserin compared to zolpidem, and 3) to evaluate the abuse potential of zolpidem compared to placebo (test of study validity).

Secondary objectives

The secondary objective was to evaluate the safety and tolerability of flibanserin.

Reviewer's comment: This review report is only for the primary objective of the study.

2.1.2 Study design

This was a single-center, randomized, double-blind, placebo- and active-controlled, 6-way crossover study of single doses of flibanserin 100 mg, flibanserin 200 mg, flibanserin 250 mg, zolpidem 15 mg, zolpidem 30 mg, and placebo in recreational polydrug users to assess the potential abuse liability of flibanserin. The study included a screening visit, a Qualification Phase, a Treatment Phase, and a post treatment follow-up/end of study visit.

Subjects were initially screened for study eligibility at a screening visit. Those who satisfied the study entry criteria during initial screening participated in a double-blind, randomized, 2-day Qualification Phase (conducted 2 to 28 days after the screening visit) during which they received a single dose of 20 mg of zolpidem and a single dose of matching placebo in a crossover manner, each of which was followed by serial pharmacodynamic assessments. Subjects who were able to discriminate between zolpidem and placebo on the pharmacodynamic assessment scales were entered into the Treatment Phase of the study, which occurred at least 7 days, but no more than 28 days, after the Qualification Phase.

Subjects were admitted to the clinical research unit (CRU) on the day before the first dose of study drug was to be administered (Day -1) and remained in the CRU for 19 consecutive nights. In Periods 1 through 6 of the Treatment Phase, flibanserin 100 mg, flibanserin 200 mg, flibanserin 250 mg, zolpidem 15 mg, zolpidem 30 mg, or placebo was administered once in a randomized, full crossover manner; each treatment was separated by a washout period of at least 48 hours. Subjects were discharged from the CRU after all of the assessments after the sixth period (last dose of study drug) had been completed. Subjects returned for a post treatment follow-up/end of study visit 5 to 14 days after the last dose of study drug.

2.1.3 Primary and secondary endpoints

The following pharmacodynamic assessments were administered to evaluate the subjective and objective effects of flibanserin.

Primary endpoint

The primary endpoint was the maximum effect (Emax) over 24 hours postdose for the Bipolar Drug Liking VAS ("at the moment") during each treatment period.

Secondary endpoints

- Balance of effects:
  - Overall Drug Liking VAS (Emax over 24 hours and at the 12- and 24-hour
  - postdose assessments during each treatment period)
  - Take Drug Again VAS (Emax over 24 hours and at the 12- and 24-hour postdose
  - Subjective Drug Value (Emax at 24 hours postdose and at the 12- and 24-hour postdose assessments during each treatment period)

- Positive effects:
  - High VAS (Emax over 24 hours postdose during each treatment period)
  - Good Drug Effects (Emax over 24 hours postdose during each treatment period)
  - ARCI Morphine Benzedrine Group (ARCI-MBG) scale (Emax over 24 hours
  - postdose during each treatment period)
    ARCI Amphetamine (ARCI-AMP) scale (Emax over 24 hours postdose during each treatment period)

- Negative effects:
  - Bad Drug Effects VAS (Emax over 24 hours postdose during each treatment
  - period)
    ARCI Lysergic Acid Diethylamide (ARCI-LSD) scale (Emax over 24 hours postdose during each treatment period)

- Sedative effects:
  - ARCI Pentobarbital, Chlorpromazine, Alcohol Group (ARCI-PCAG) scale (Emax over 24 hours postdose during each treatment period)
  - Alertness/Drowsiness VAS (Emax [alertness] and Emin [drowsiness] over 24 hours postdose during each treatment period) and time to maximum effect (Tmax)

- Other drug effects:
  - Any Drug Effects VAS (Emax over 24 hours postdose during each treatment period)
  - Drug Similarity VAS (arithmetic mean score at 12 hours postdose during each treatment period)

Reviewer's comments: Emax was the maximum of postdose responses during 8 hours in the reviewer's analysis.

2.1.4 Number of subjects

A total of 106 subjects entered the Qualification Phase of whom 104 subjects (98.1%) received both the single dose of placebo and the single dose of 20 mg of zolpidem and completed all of the Qualification Phase study activities. Of the 104 subjects who completed the Qualification Phase, 36 subjects entered the Treatment Phase and were randomized to 1 of the 6 treatment sequences; 34 of the 36 subjects (94.4%) received all 6 doses of study drug (flibanserin 100 mg, flibanserin 200 mg, flibanserin 250 mg, zolpidem 15 mg, zolpidem 30 mg, and placebo), completed all study activities, and were included in the pharmacodynamic analyses. All 36 of the subjects who entered the Treatment Phase were included in the safety analyses.

2.1.5 Statistical methodologies used in the Sponsor's analyses

All pharmacodynamic results were summarized in tabular format, with summary statistics (number[n], mean, standard deviation [SD], median, minimum, and maximum) and change from baseline, as appropriate by time point.

The primary and secondary endpoints for the completer population were analyzed using a standard mixed-effect analysis of variance (ANOVA) model for a crossover study. Sequence, subject (sequence), period, and treatment were incorporated into the model. Baseline (predose Period 1/Day 1) measurement, as a covariate, was included in the model where applicable. If the p-value of the baseline*treatment interaction term was > 0.01, then the interaction term was removed from the model (Reduced Model) and the analysis was rerun on the Reduced Model. If the baseline*treatment interaction was statistically significant and not due to an outlier, least squares (LS) treatment means for change from baseline were displayed with baseline set to the 25th quartile, the median, the mean, and the 75th quartile.

An analysis of the normality of the data distribution for the primary endpoint was conducted on the residuals of the model using the Shapiro-Wilk Test at alpha = 0.01. As the test for normality was not statistically significant for the primary endpoint, no nonparametric analyses were conducted.

Only summary statistics were provided for secondary endpoints for which a subset of the completer population answered the specified question. In addition, exploratory analyses of any outlier responses (e.g., analysis for the completer population and analysis with outlier subject removed from the completer population) were conducted as appropriate.

2.1.5 Sponsor's Conclusions

The findings that all 3 doses of flibanserin were associated with significantly less "at the moment" drug liking than the 30-mg dose of zolpidem, coupled with its sedative effects and the fact that flibanserin was generally associated with fewer positive effects than the 30-mg dose of zolpidem and the "subjective drug value" for flibanserin was generally less than for zolpidem, indicate that flibanserin, even at a dose of 250 mg (2.5 times the proposed therapeutic dose), is less preferred than a 30-mg dose of zolpidem. Single doses of 100, 200, and 250 mg of flibanserin and single doses of 15 and 30 mg of zolpidem were well tolerated in this population of recreational polydrug users, with no new safety concerns identified.

2.2 Data Location

The analysis datasets are located at

\\cdsesub1\evsprod\nda022526\0042\m5\datasets\spr-12-05\analysis\legacy\datasets\adqsde.xpt
\\cdsesub1\evsprod\nda022526\0042\m5\datasets\spr-12-05\analysis\legacy\datasets\adqsar.xpt
\\cdsesub1\evsprod\nda022526\0042\m5\datasets\spr-12-05\analysis\legacy\datasets\adqshl.xpt
\\cdsesub1\evsprod\nda022526\0042\m5\datasets\spr-12-05\analysis\legacy\datasets\adqssd.xpt

2.3 Reviewer's Assessment

In the reviewer's report P, Z15, Z30, F100, F200, and F250 denote placebo, zolpidem 15 mg and 30 mg, flibanserin 100 mg, 200 mg, and 250 mg, respectively.

2.3.1 Primary Analysis

2.3.1.1 Descriptive Statistics

Table 1 summarizes the mean, standard error, minimum, the first quartile (Q_1), median, the third quartile (Q_3), and maximum for six treatments in the study and for the treatment differences between flibanserin and zolpidem or placebo, as well as among doses of each drug for Emax of Drug Liking VAS.

Table 1: Summary statistics for Emax of Drug Liking VAS (N=34)

TRT or Comparison	Mean	StdErr	Min	Q1	Med	Q3	Max
F100	61.12	2.25	47	51.0	55.0	68.5	96
F200	66.68	2.78	46	51.0	69.5	76.3	99
F250	66.32	2.62	47	53.0	66.5	77.3	98
P	53.82	1.03	48	51.0	52.0	53.3	70
Z15	68.53	2.51	48	52.8	70.0	79.3	100
Z30	74.06	2.88	50	56.8	74.0	87.3	100
F100-P	7.29	2.39	-17	-0.3	2.0	14.0	46
F100-Z15	-7.41	2.39	-41	-14.8	-4.0	1.0	19
F100-Z30	-12.94	2.51	-49	-20.8	-12.0	-1.0	6
F200-P	12.85	3.02	-17	0.0	3.5	26.3	48
F200-Z15	-1.85	2.24	-41	-5.5	-0.5	6.0	21
F200-Z30	-7.38	2.70	-49	-14.0	-6.0	-0.8	42
F250-P	12.50	2.83	-16	0.0	4.5	26.3	48
F250-Z15	-2.21	2.05	-44	-6.5	-2.0	4.5	18
F250-Z30	-7.74	3.10	-52	-16.0	-7.5	-0.5	41
Z15-P	14.71	2.67	-18	0.8	12.5	27.5	48
Z30-P	20.24	3.11	-20	2.8	20.5	35.0	51
Z30-Z15	5.53	2.61	-43	1.8	4.5	15.0	37
F200-F100	5.56	2.03	-16	-1.3	2.0	13.5	36
F250-F100	5.21	2.18	-18	-0.3	3.0	14.3	35
F250-F200	-0.35	1.86	-26	-3.0	0.0	2.3	33

As seen in Table 1, the first quartiles of Z15 and Z30 are 52.8 and 56.8 respectively. These values are within the range of neutral response (~40 to 60) for a bipolar scale such as Drug Liking VAS. This suggests that approximate 25% of subjects (or possibly more) did not respond to the positive control in this study.

Figure 1 provides the boxplots of six treatments as well as boxplots for the differences in Emaxs between flibanserin and zolpidem, and between flibanserin and placebo for Drug Liking VAS. The line in each box denotes the median, the circle in each box is for the mean, and the plus sign indicates that the mean approximately equals the median. The blue presents the neutral score 50, and the red line presents strongly dislike or 0 difference in responses between two treatments.

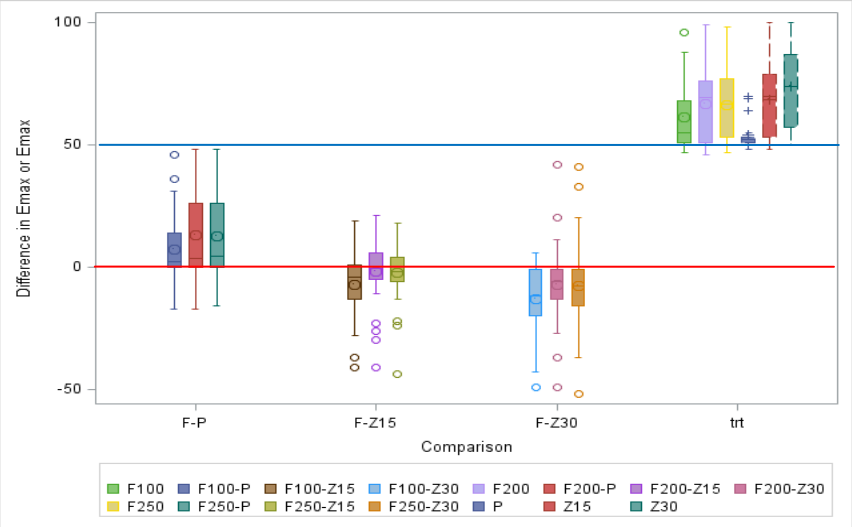


Figure 1: Boxplots for six treatments and the differences between flibanserin and zolpidem, and between flibanserin and placebo for Drug Liking VAS (N=34)

Figure 2 plots the mean dose response curves based on the least square means for flibanserin and zolpidem for Drug Liking VAS.

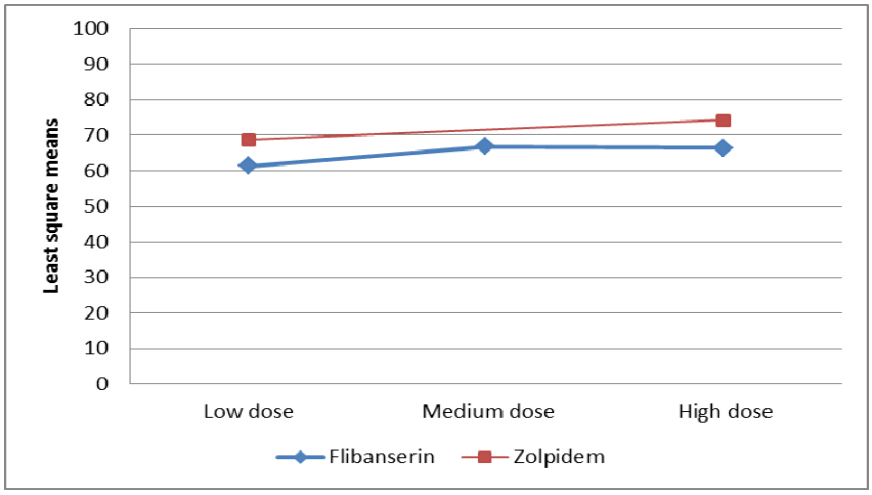


Figure 2: Mean dose response curves for Drug Liking VAS (N=34)

Figure 3 is the mean time course profiles for Drug Liking VAS. Three profiles from flibanserin are under those of zolpidem. The profiles of F100 and F250 are around neutral score 50 in the early hours, and then are below 50, especially for F250. The peak mean response to F200 is around 57 at hour 1, while the peak mean responses to Z15 and Z30 are approximately 62 and 64 at hour 1 and 1.5, respectively.

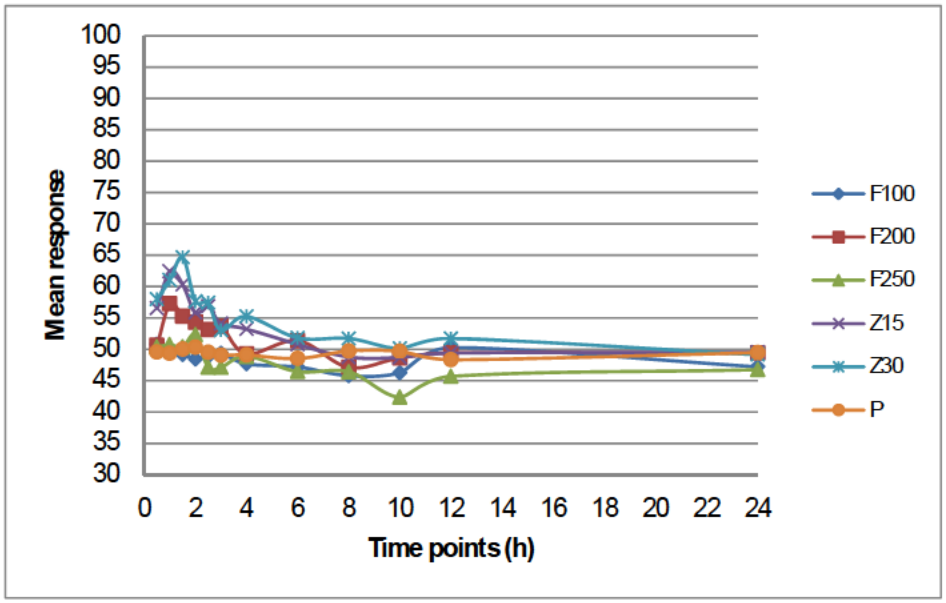


Figure 3: Mean time course profiles for Drug Liking VAS (N=34)

Figure 4 is a heat map display, which shows individual Emax responses to each treatment as measured by Drug Liking VAS. In the figure, drug disliking is shown in blue, neutral response is shown in white and drug liking is shown in pink. The subjects (n = 34) represented above the thin orange line are females (n = 9), and the subjects below the orange line are males (n = 25).

Figure 1 shows that 5 of 34 subjects (14.7%) of subjects had an Emax to placebo greater than 60 for Drug Liking VAS, which is outside the acceptable neutral response for placebo that ranges from 40-60. In contrast, 11 of 34 subjects (32.6%) who received Z15 and 9 of 34 subjects (26.5%) show received Z30 had an Emax to placebo within the neutral range. The actual Emaxs for these subjects were between 48 and 53 and between 50 and 57 for Z15 and Z30, respectively.

Figures 5 and 6 are the individual time course response profiles for Z15 and Z30, respectively. One may see from the graphics the responses from all subjects at each time point, and the peak response and the duration of the peak response for each subject to the positive control.

Using Figures 4-6, one can easily locate the subjects who did not respond to the positive control.

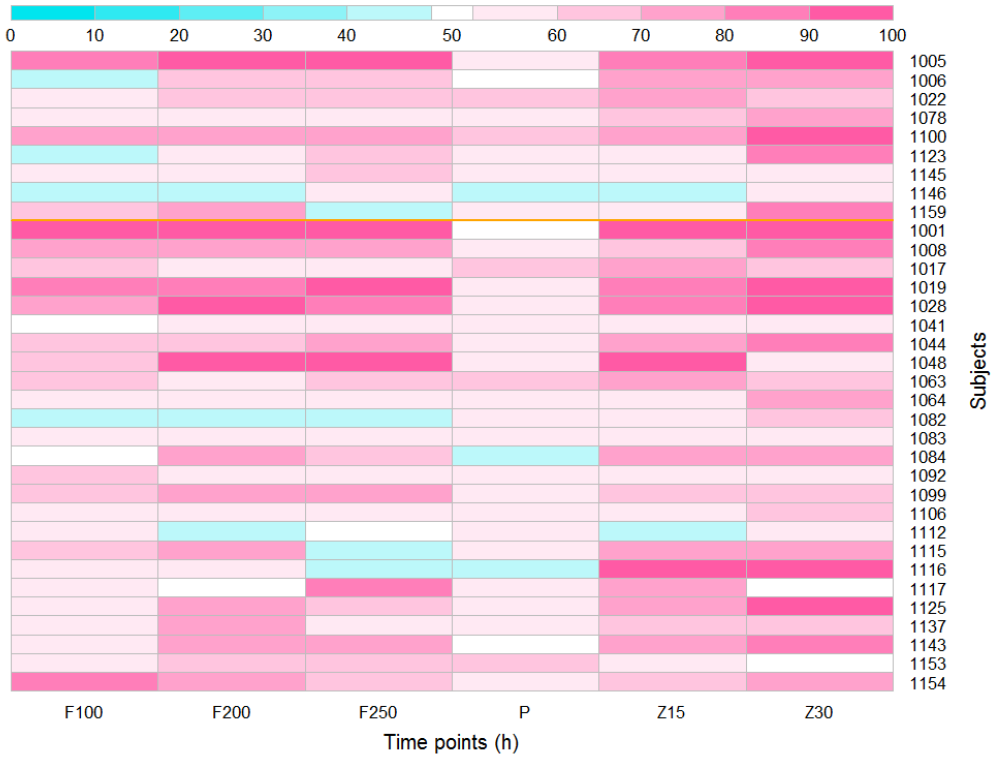


Figure 4: Heat map for Emax of Drug Liking VAS by treatment

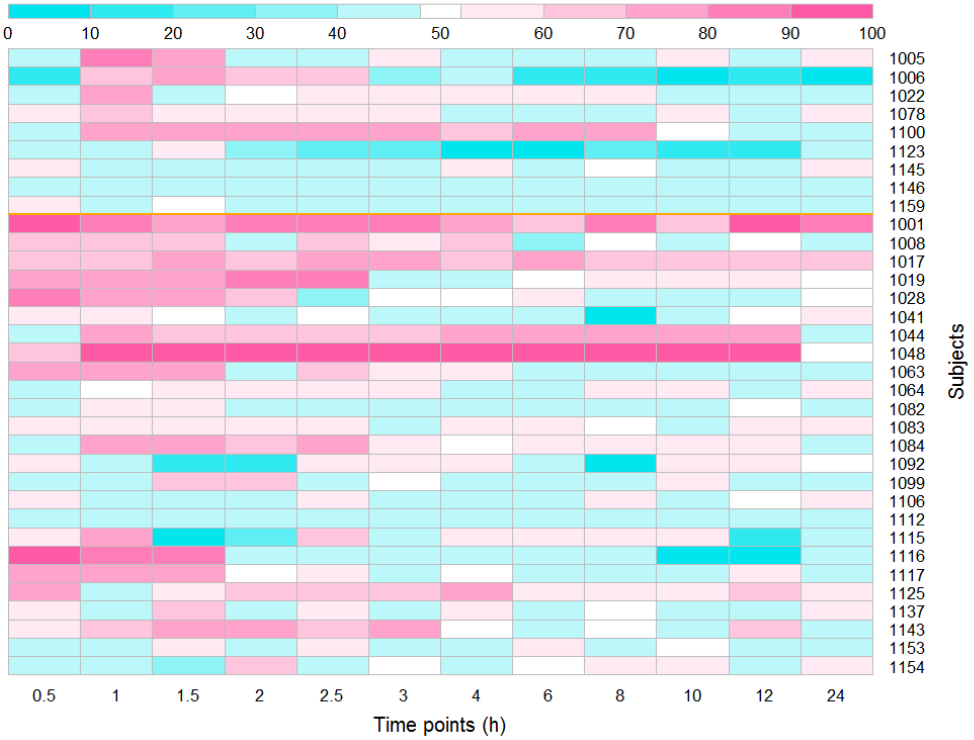


Figure 5: Individual time course response profiles for Drug Liking VAS (Z15)

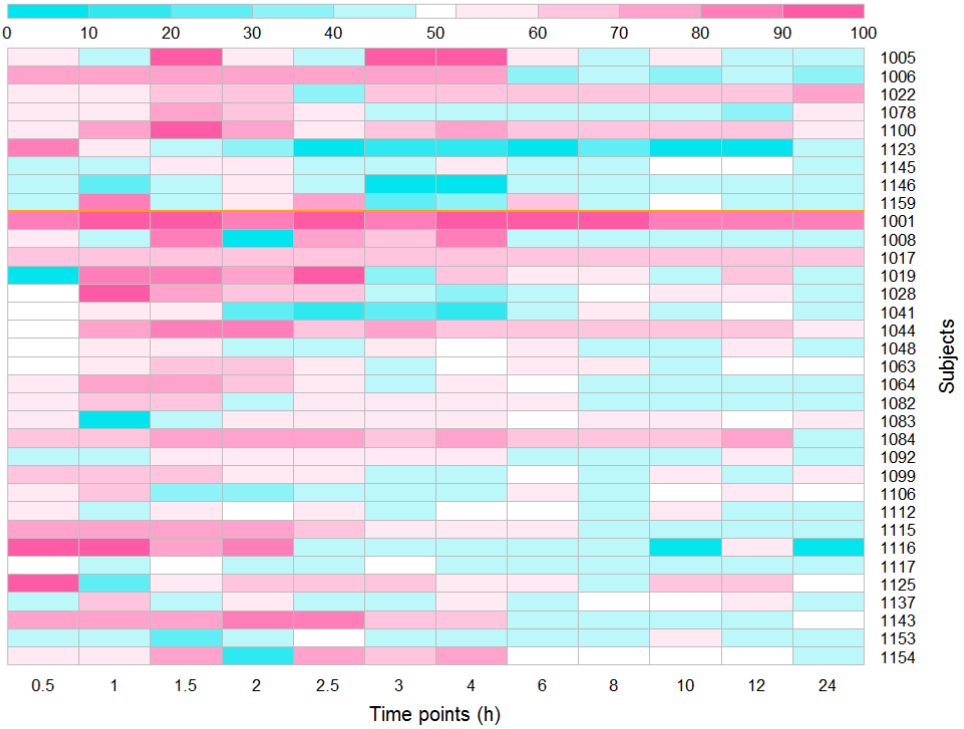


Figure 6: Individual time course response profiles for Drug Liking VAS (Z30)

2.3.1.2 Statistical Testing

The statistical model used in the reviewer's analysis was the mixed-effects model with period, sequence, and treatment as fixed effects, and subject as a random effect. The reviewer checked assumptions in the model for the equal variances and the normality. The normal assumption was not violated for Drug Liking VAS. However, the assumption of equal variances was not satisfied. The SAS proc mixed procedure can adjust the unequal variances using Tukey-Kramer's method.

Table 2 shows the primary analysis results. The least square mean and 95% confidence interval for the mean of each treatment are shown in the third row of Table 2. Rows 4-5 show the mean differences and p-values for the comparisons between flibanserin and each dose of zolpidem. The last row is for the comparisons between each dose of flibanserin and placebo as well as each dose of zolpidem and placebo.

Table 2: Statistical analysis results for Drug Liking VAS

Measure	Treatment	F100	F200	F250	Z15	Z30	P
	N	34	34	34	34	34	34
Drug Liking VAS	LSmean 95% CI	61.37 (56.50, 66.25)	66.78 (61.91, 71.65)	66.36 (61.49, 71.23)	68.79 (63.91, 73.66)	74.16 (69.29, 79.04)	53.86 (48.99, 58.73)
	Diff vs Z15/ pval	-7.41 / 0.0423	-2.00 / 0.9678	-2.43 / 0.9284			
	Diff vs Z30/ pval	-12.79 / <.0001	-7.38 / 0.0437	-7.81 / 0.0272			
	Diff vs P/ pval	7.52 / 0.0379	12.92 / <.0001	12.50 / <.0001	14.93 / <.0001	20.31 / <.0001	

Note: pval denotes p-value. All p-values were from the two-sided t test, and adjusted by Tukey -Kramer's method for unequal variances.

The primary analysis shows that

1. For each dose of flibanserin (F100, F200 and F250), on the average, there was a statistically significant increase in response on Drug Liking VAS compared to placebo ($p = 0.04, < 0.0001$ and < 0.0001, respectively).
2. On the average, there was no significant difference in responses between two high doses of flibanserin (F200 and F250) and Z15 on Drug Liking VAS ($p >> 0.05$).
3. On the average, all doses of flibanserin produce positive subjective responses that were statistically lower than Z30 ($p < 0.0001$, 0.0437, and 0.0272).
4. The validation tests (compared zolpidem 15 mg and 30 mg to placebo) were statistically significant ($p<0.0001$).

However, notice that more than 25% of subjects did not respond to the positive control, and twenty five percent of appropriately selected subjects could change the results from the primary analysis.

2.3.2 Secondary Analysis

After examining the data for the secondary measures, this reviewer found the following problems in this study.

1. Large placebo responses were observed for many secondary measures in the study. For example: 14 of 34 subjects (41.2%) had Emax of Good Effects VAS to placebo greater than or equal to 39 on the unipolar scale from 0 to 100.

   Figures 7-10 show placebo responses from individual subjects for Good Effects, High VAS, ARCI MBG and ARCI AMP, respectively.

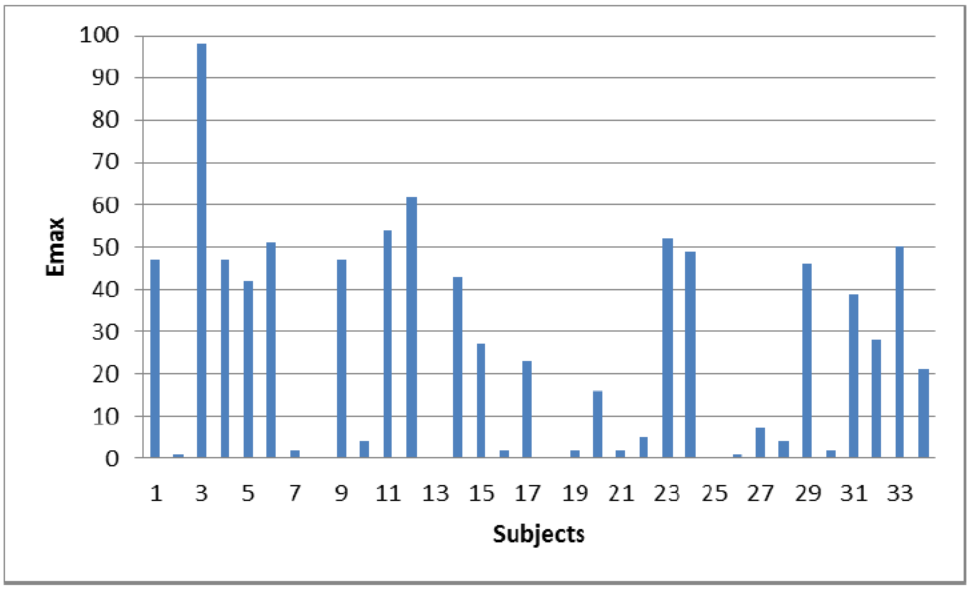


Figure 7: Placebo responses from individual subjects for Good Effects VAS

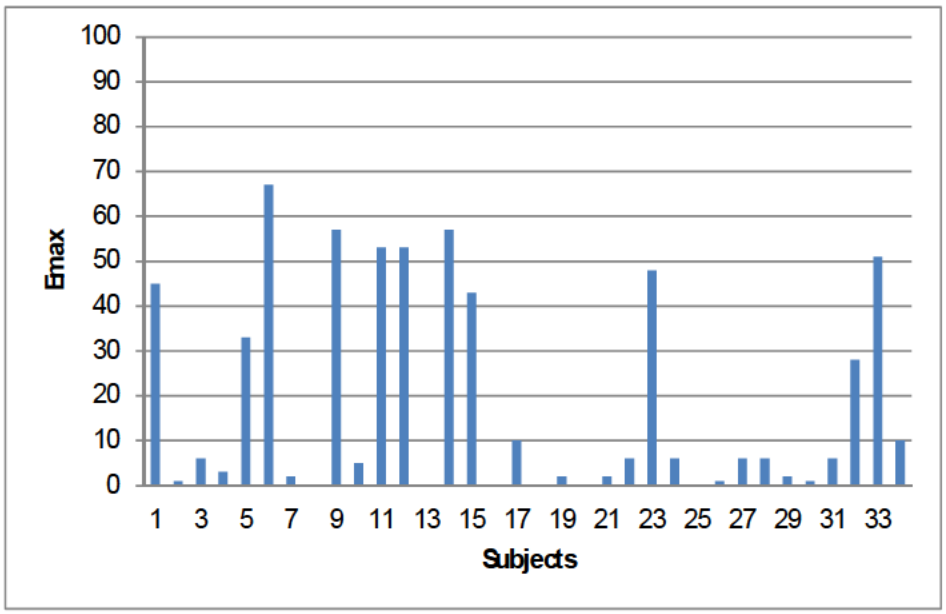


Figure 8: Placebo responses from individual subjects for High VAS

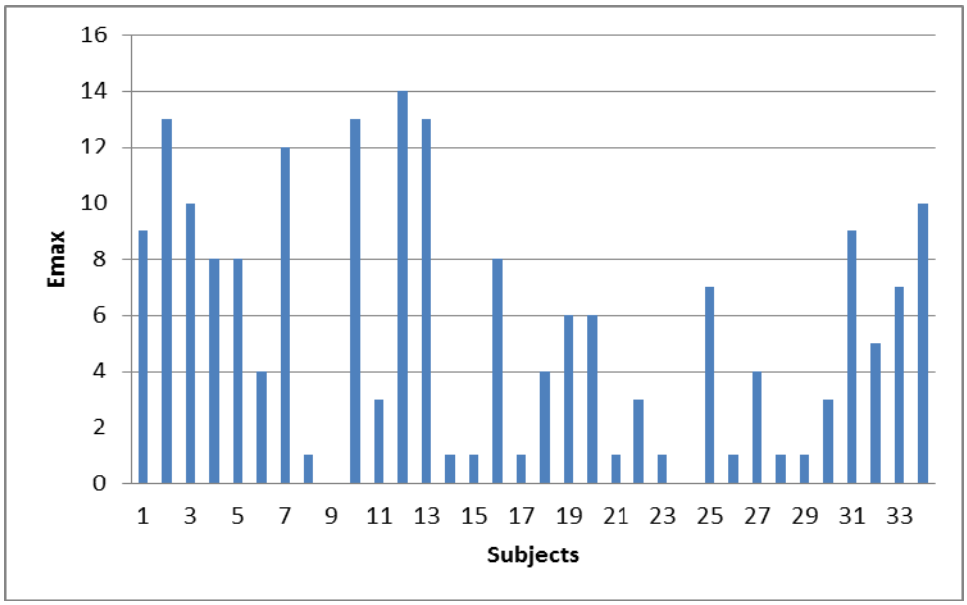


Figure 9: Placebo responses from individual subjects for ARCI MBG

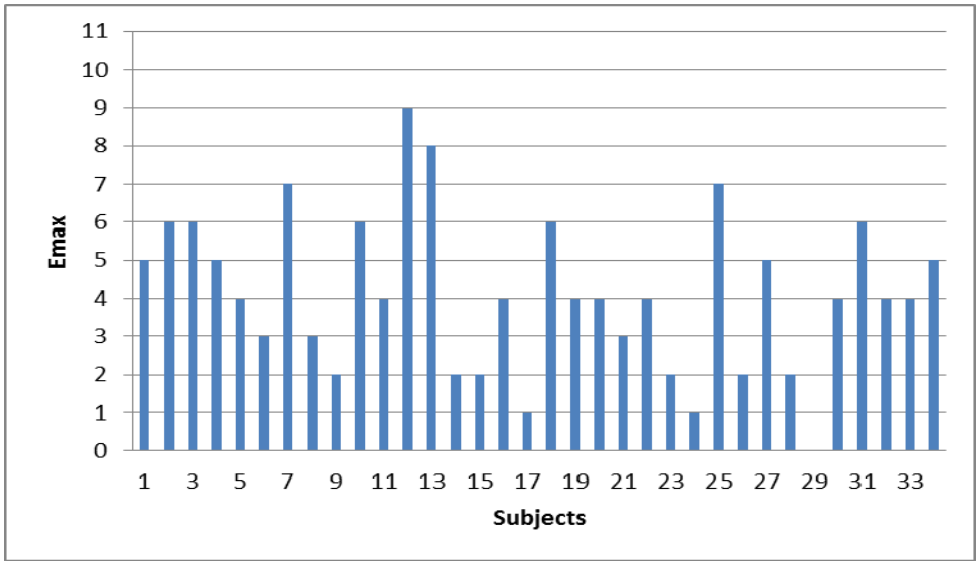


Figure 10: Placebo responses from individual subjects for ARCI AMP

Subjects might be confused by unipolar and bipolar visual analog scales. However, large placebo responses are also observed for ARCI MBG and ARCI AMP, which are not on a visual analog scale.

2. For some abuse potential measures, such as High VAS and ARCI measures, predose responses were collected in each treatment period before dosing for each subject in the study. However, these predose responses were not utilized in the statistical analysis. Instead, the Sponsor inappropriately used baseline responses (defined as predose

responses for each subject in Period 1 on Day 1) in calculating the response variable and the covariate in the statistical model for the analysis.

3. Large predose responses were also observed in the study. Figures 11 and 12 are two examples for High VAS on a unipolar scale ranged from 0 to 100 and ARCI MBG on a scale ranged from 0 to 16.

Figure 11 shows the predose responses from individual subjects in each treatment period for High VAS. From the graph, one may see the predose responses from subjects in each treatment period, and the predose responses from each subjects in all treatment periods. The color represents the predose response score. For example, 32 of 34 subjects (94.1%) had predose responses between 40 and 60 for High VAS in the first treatment period. From the data, we know that these predose responses are, in fact, between 46 and 51 in the first treatment period. Because these predose responses were so close to 50, it appears that subjects were confused by unipolar scale or bipolar scale.

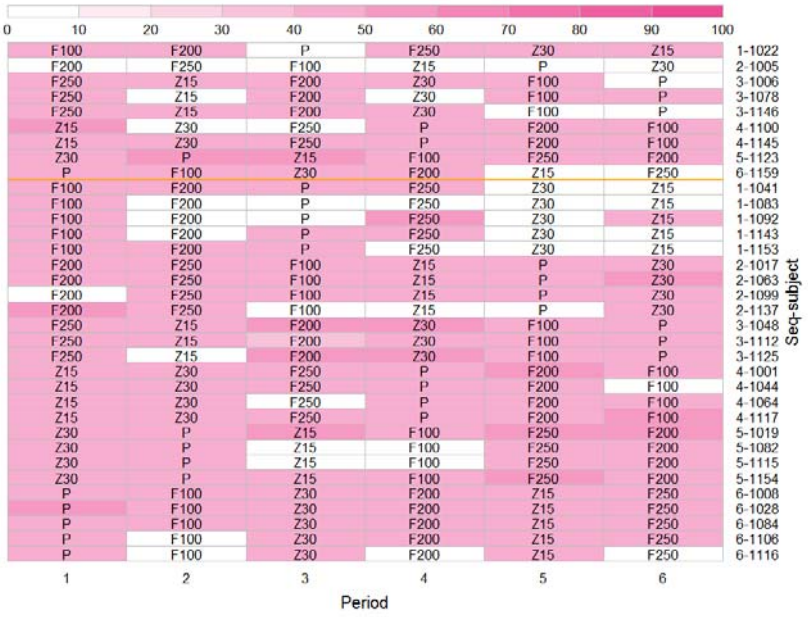


Figure 11: Predose responses from individual subjects in each treatment period for High VAS

Figure 12 shows the predose responses from individual subjects in each treatment period for ARCI MBG. More than 25% of subjects had consistently moderate to large predose responses in all treatment period. For example, in the first treatment period, 12 of 34 subjects (35%) had predose responses for ARCI MBG between 7 and 14. Among these subjects, 6 subjects had score 10 or above for this measure in the first.

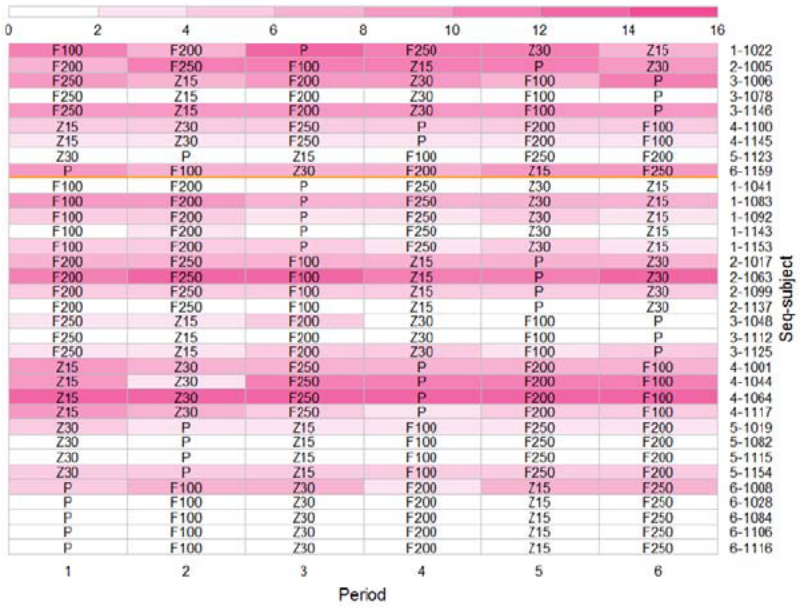


Figure 12: Predose responses from individual subjects in each treatment period for ARCI MBG

4. Overall Drug Liking VAS and Take Drug Again VAS are on a bipolar scale from 0 to 100. The neutral score is 50. In this study, for Overall Drug Liking VAS 25 of 34 subjects (73.5%) and 24 of 34 subjects (70.6%) had an Emax less than 55 to zolpidem 15 mg and 30 mg, respectively, and for Take Drug Again VAS 22 of 34 subjects (64.7%) and 21 of 34 subjects (61.8%) had an Emax less than 55 for zolpidem 15 mg and 30 mg, respectively. Figures 13 and14 show the Emaxs (the maximum response at hours 12 and 24) of Overall Drug Liking VAS and Take Drug Again from each subject to Z15 and Z30, respectively. One may see from these graphs that many subjects had scores around 50.

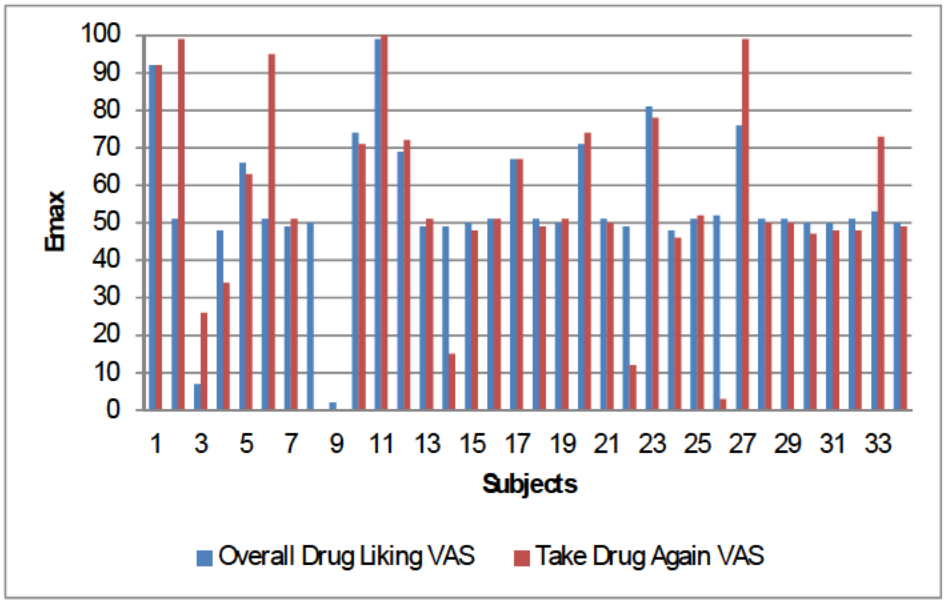


Figure 13: Emaxs from individual subjects to Z15 for Overall Drug Liking VAS and Take Drug Again VAS

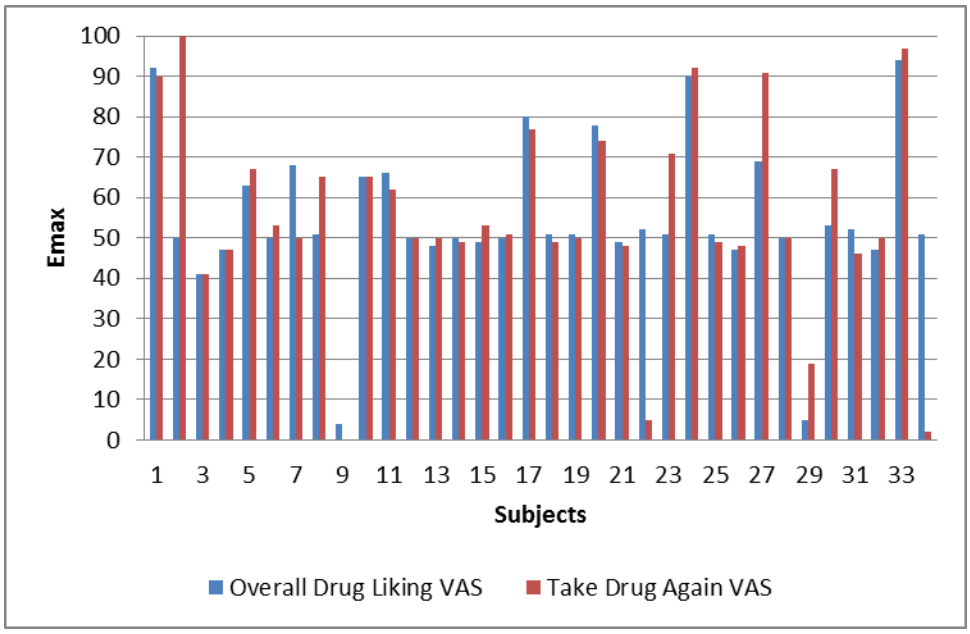


Figure 14: Emaxs from individual subjects to Z30 for Overall Drug Liking VAS and Take Drug Again VAS

3. Conclusion

The findings from this review for Study SPR12-05 are summarized below:

1. More than 25% of subjects did not respond to the positive control for the primary measure Drug Liking VAS.
2. Large placebo responses were observed for many secondary measures.
3. The predose responses were not used in the statistical analysis. Instead, the Sponsor inappropriately used baseline responses (defined as predose responses for each subject in Period 1 on Day 1) in calculating the response variable and the covariate in the statistical model for the analysis.
4. Many large predose responses were observed for Good Effects VAS, High VAS, ARCI MBG and ARCI AMP.
5. Majority of subjects had Emax around 50 or below to the positive control for Overall Drug Liking VAS and Take Drug Again VAS.

In conclusion, the subjects were not well selected and well trained. With the problems stated in previous section, the results from this study are not interpretable. Therefore, the study is inconclusive.

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

/s/

LING CHEN
08/20/2013

YI TSONG
08/21/2013

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

Statistical Review and Evaluation

CLINICAL STUDIES

NDA:	NDA 22-526
Drug Name:	Girosa (flibanserin) tablets, 100mg
Indication(s):	Treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women
Applicant:	Boehringer Ingelheim Pharmaceuticals, Inc.
Date(s):	Letter date: October 27, 2009
	PDUFA goal date: August 26, 2010
Review Priority:	Standard
Biometrics Division:	DB3
Statistical Reviewer:	Lisa A. Kammerman, Ph.D.
Concurring Reviewers:	Mahboob Sobhan, Ph.D.
Medical Division:	Division of Reproductive and Urologic Drug Products
Clinical Team:	Daniel Davis, MD Olivia Easley, MD Lisa Soule, MD
Project Manager:	Charlene Williamson

Keywords: Clinical studies, NDA review, patient-reported outcomes, missing data

Table of Contents

FIGURES

TABLES

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

From a statistical perspective, the evidence submitted with this application does not support the efficacy of flibanserin 100 mg q.h.s. for the treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women. I base my conclusion primarily on the finding that the two major Phase 3 studies, Study 511.71 and Study 511.75, failed to achieve statistical significance on both pre-specified co-primary endpoints. The treatment difference between flibanserin 100 mg q.h.s. and placebo was statistically significant in both studies for sexually satisfying events. However, the treatment difference for sexual desire as measured by the daily electronic diary (eDiary Desire) was not statistically significant in either study.

I recommend exploratory analyses, to be done by the Applicant, of the relationship between the baseline value of eDiary Desire and the difference between treatments at 24 weeks. If the results of these exploratory analyses suggest the possibility of a treatment effect among subjects who have no or minimal desire, the Applicant may wish to consider limiting enrollment to these subjects in future studies of flibanserin for the treatment of HSDD. This recommendation is based on my preliminary exploratory analyses of eDiary Desire, which suggest the difference at 24 weeks between flibanserin 100 mg q.h.s. and placebo depended on the baseline value of eDiary Desire. The observed treatment difference was greatest among subjects whose sexual desire value at baseline was relatively low. The observed treatment difference decreased with increasing values of sexual desire at baseline.

1.2 Brief Overview of Clinical Studies

The submission contains five randomized, double-blind, placebo-controlled studies and two open-label studies. The studies include a variety of flibanserin regimens; see Table 1.

To support the approval of flibanserin 100 mg q.h.s, the submission emphasizes the results from the three double-blind North American studies that evaluated the 100 mg q.h.s regimen:

- 511.71 – A 24-week, randomized, double-blind, placebo-controlled, safety and efficacy trial of flibanserin 50 and 100 milligrams each evening in premenopausal women with Hypoactive Sexual Desire Disorder.
- 511.75 – A 24-week, randomized, double-blind, placebo-controlled, safety and efficacy trial of flibanserin 50 milligrams and, with up titration, 100 milligrams daily in premenopausal women with Hypoactive Sexual Desire Disorder.
- 511.74 – A 48-week, randomized discontinuation trial of flibanserin in women with hypoactive sexual desire disorder containing an open-label, flexible dose period followed by a double-blind, randomized, placebo-controlled period.

The purpose of Study 511.74 was to assess maintenance of efficacy and to assess potential withdrawal effects. The study was designed as a randomized withdrawal study in which patients first completed a 6-month open-label (OL) flibanserin treatment period after which responders entered a 6-month double-blind, placebo-controlled, randomized withdrawal period.

A fourth study that was conducted in North America, Study 511.70, did not evaluate the flibanserin regimen (100 mg q.h.s) for which the applicant is seeking approval.

Although Study 511.77 was double-blind and studied the dosage of interest, the NDA designates this study as "supportive" because it was conducted in Europe.

Two additions studies (511.84 and 511.118) were open-label, uncontrolled extensions of the North American and European studies.

Table 1. Overview of Efficacy and Safety Studies

BI Trial Number		Placebo	FLI 25 mg b.i.d.	FLI 50 mg q.h.s.	FLI 50 mg b.i.d.	FLI 100 mg q.h.s.
North American trials						
Phase III: 24-week, placebo-controlled, parallel group, double-blind						
511.70		X	X	X	X[1]	
511.71		X		X		X
511.75		X	X		X[1]	X[1]
Phase III: Randomized withdrawal – 24-week, open-label, flexible flibanserin regimen, followed by 24-week, placebo-controlled, double-blind, randomized withdrawal						
511.74	Open-label			X[2]	X[3]	X[3]
	Double-blind	X		X	X	X
Phase III: Long-term exposure						
511.84			X[4]	X[4]	X[4]	X[4]
European trials						
511.77		X		X		X[1]
511.118			X[4]	X[4]	X[4]	X[4]

1 Up-titration from FLI 50 mg q.h.s. after the first 2 weeks of treatment as part of the dosing regimen
2 Patients start on FLI 50 mg q.h.s.
3 Optional up-titration after at least 4 weeks of FLI 50 mg q.h.s. treatment
4 Dose available as part of a flexible dosing arm
Source: Summary of Clinical Efficacy [Module 2.7.3], Table 1.1: 2

Source: Table 2.5.6.2.1:1 of Clinical Overview

The entry criteria for Study 511.71, Study 511.74, Study 511.75 and Study 511.77 included:

- Premenopausal women
- 18 years of age and older
- Primary diagnosis of HSDD, generalized acquired type, according to DSM-IV-TR criteria. The current episode must be at least 24 weeks in duration by the baseline visit.
- A score of "15" or higher on the Female Sexual Distress Scale-Revised (FSDS-R) at

the Screen Visit

An additional criterion for Study 511.71, Study 511.74 and Study 511.77 was:

- Item Number Two of the Sexual Interest and Desire Inventory – Female (SIDI-F) must be rated as "0" or "1" at the Screen Visit

These four studies excluded women with other diagnoses of sexual disorder or dysfunction, major depressive disorder or disorders that could affect sexual function, and women who started psychotherapeutic (non-drug) treatment within 12 weeks of the baseline visit.

1.3 Statistical Issues and Findings

Only Study 511.74 – the randomized, withdrawal study – was a positive study; see Table 2. None of the other studies won on both co-primary endpoints. Studies 511.71 and 511.75 were statistically significant for satisfying sexual events (SSE) but not sexual desire. Study 511.77 did not yield a statistically significant result for SSE; therefore, testing of other endpoints was not appropriate.

Section 2.1.3 discusses the major design and statistical issues associated with the two main Phase 3 studies, Study 511.71 and 511.75. Briefly, these issues include:

- Only one prespecified co-primary endpoint was statistically significant (satisfying sexual events).
- Study discontinuations due to non-compliance were stipulated by the study protocols, resulting in a high rate of missing data.
- Missing data on the eDiary desire endpoint are informative.
- Applicant's request to replace the eDiary sexual desire co-primary endpoint with FSFI-desire, a pre-specified secondary endpoint, resulting in statistical significance
- 28-day windows were anchored to clinic visits, rather than using available diary data divided into 28-day windows starting with the date of randomization

The following paragraphs provide an overview and summary of some of the key design and statistical issues that need to be considered when interpreting the study results and evaluating the Applicant's conclusions regarding the efficacy of flibanserin.

The application indicates the eDiary measure of sexual desire was not adequately validated and suggests the lack of validation along with the frequency of noncompliance for this item may partially explain its non-significant results in the two studies. Noncompliance for the eDiary sexual desire endpoint, however, appears to provide information on a subject's level of desire and should not be treated as missing data. For example, analyses show subjects who did not respond to the eDiary desire time had a mean number of sexually satisfying sexual events (SSEs) that fell between the mean number of SSEs for subjects reporting no sexual desire and those

reporting minimal sexual desire. This finding suggests that a missing value on the eDiary sexual desire item likely represents a subject whose level of desire was, at most, minimal as measured by the eDiary.

The submission proposes replacing the eDiary sexual desire item with the Female Sexual Function Index – Desire Items (FSFI-desire), which the submission claims is a better instrument for assessing sexual desire and which does not have an issue with noncompliance, because it is administered in the clinic. The Applicant's background materials for the advisory committee meeting that was held on June 18, 2010 indicate that limitations of the eDiary Desire item became evident as the clinical program progressed. Information contained in the NDA, however, is insufficient to conclude the FSFI-desire index is validated for use in a clinical trial enrolling this patient population.

If FSFI-desire is used in place of eDiary Desire as the co-primary endpoint measuring sexual desire, the results are statistically significant in both studies. From a statistical perspective, however, in addition to the lack of information supporting the validation of FSFI-desire, the use of FSFI-desire in place of the eDiary Desire endpoint is problematic, especially because FSFI-desire was not a prespecified co-primary endpoint. Although there can be situations where an endpoint other than the pre-specified endpoint could be used to assess the efficacy of a product, the substitution of a patient-reported outcome as a primary endpoint after the data are unblinded does not seem to be such a situation. In addition, the FSFI was one of a series of tests that were administered during clinic visits. The submission does not address the effects that completing these other tests may have had on a subject's responses to the FSFI and on the content validity of the FSFI.

Exploratory analyses of eDiary Desire suggest the difference at 24 weeks between flibanserin 100 mg q.h.s. and placebo depended on the baseline value of eDiary Desire. The observed treatment difference was greatest among subjects whose eDiary Desire value at baseline was relatively low. The observed treatment difference decreased with increasing values of eDiary Desire at baseline. I recommend the Applicant doing additional analyses that explore the relationship between the baseline value of eDiary Desire and the difference between treatments at 24 weeks. If these exploratory analyses suggest the possibility of a treatment effect among subjects who have no or minimal desire, the Applicant may wish to consider limiting enrollment to these subjects in future studies of flibanserin for the treatment of HSDD.

Although the protocols called for analyses of change between the 4-week baseline and the interval of Weeks 21 to 24, the analyses of the co-primary endpoints did not adhere to this plan. Instead, the eDiary data were divided into 4-week intervals that were anchored to a subject's clinic visits, and the eDiary data for the co-primary endpoints were computed as a 28-day average of all data collected between visits. As expected, clinic visits did not necessarily occur every four weeks; visits included a window of ±7 days. Moreover, the primary analyses required a minimum of 14 days of data. If less than 14 days of data were available, data were imputed by last observation carried forward. The results of analyses that did not impute for missing data,

which used all available data, were consistent with the results of the analyses that used last observation carried forward.

Using 28-day intervals that started with the day of randomization, the Applicant repeated the analyses of the co-primary endpoints, both for the endpoints standardized to 28 days and for the actual values observed. For the SSE endpoint, the results are not as strong as the results reported for the windows anchored to clinic visits. Although the results for the data standardized to 28 days are consistent with the original analyses, the treatment difference for SSEs in Study 511.75 moves closer to statistical non-significance (p=0.048, adjusted for multiplicity by Hochberg's method). In the analyses of all available data, the treatment difference for SSEs is smaller and is statistically non-significant in Study 511.75.

2 INTRODUCTION

2.1 Overview

The Applicant is seeking approval of flibanserin 100 mg q.h.s for the treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women. Currently, there are no products approved for the treatment of HSDD. Flibanserin would be the first product to be marketed for this indication.

To support the indication, the Applicant submitted numerous studies. Section 2.1.1 identifies the studies that are the focus of my statistical review. Section 2.1.2 presents major design and statistical issues that need to be considered when interpreting the results of the studies.

2.1.1 Studies selected for statistical review

The Applicant submitted four Phase 3 studies that evaluated the 100 mg q.h.s regimen; see Table 1. My review focuses on two studies, Studies 511.71 and 511.75, which the Applicant is using to support their claim that flibanserin 100 mg q.h.s is effective for the treatment of HSDD. Both were double-blind, placebo-controlled, randomized studies conducted in North America.

The third study, Study 511.74, may have provided insight into the maintenance of the treatment effect had Studies 511.71 and 511.75 provided convincing evidence of efficacy. Study 511.74 was a randomized withdrawal study in which subjects first completed a 6-month open label flibanserin treatment period, after which responders entered a 6-month double-blind, placebo-controlled, randomized withdrawal period. Because of the study's long run-in period and the highly selective sample of subjects who were randomized to treatment or placebo, this study is of limited help in establishing the efficacy of flibanserin for the treatment of HSDD.

The fourth study, Study 511.77, was double-blind and studied the 100 mg regimen. The results from this study were statistically non-significant. The Applicant deems this study as "supportive" because it was conducted in Europe – not North America.

The other studies either did not study flibanserin 100 mg q.h.s or were open-label.

2.1.2 Co-primary endpoints

The protocols for Study 511.71 and Study 511.75 identified two co-primary endpoints, each of which is a patient-reported outcome[1]:

- Frequency of satisfactory sexual events (SSEs), and
- Sum of responses for sexual desire

The primary statistical analyses assessed:

- Change in the number of satisfactory sexual events (SSEs) from the four-week baseline to the 28 days prior to the final clinic visit (without spanning into the previous clinic visit), and
- Change in the eDiary sexual desire score from the four-week baseline to the 28 days prior to the final clinic visit (without spanning into the previous clinic visit)

In order for a clinical study to have a successful outcome, the study needed to demonstrate both statistical and clinical significance for treatment versus placebo for each of the co-primary endpoints.

Study subjects used a personal handheld electronic device (eDiary) to record sexual activity, including sexual encounters and orgasms, desire and distress; see Appendix I for the questions and responses that were posed to subjects. The eDiary information was to be recorded daily and the data were to be transferred to a vendor on a daily basis by a modem.

Sexual events or encounters included sexual intercourse, oral sex, masturbation or genital stimulation by the partner. The subject herself judged whether the event was satisfactory by answering the question: "was the sex satisfying for you?"

Sexual desire was assessed by asking the subject to indicate her most intense level of sexual desire (No desire, Low desire, Moderate desire, Strong desire).

If a subject missed a day, the eDiary did not ask her to enter desire or distress information for the missed 24 hours and that data were considered missing. However, this was not the case for sexual activity. The eDiary prompted a subject to enter information on sexual activity retrospectively for up to 72 hours. If a subject missed more than three consecutive days, sexual activity beyond the last 72 hours was considered missing.
The number of SSEs in the 28 days prior to the final clinic visit and the eDiary sexual desire

[1] The medical division and the Applicant discussed these endpoints at the End-of-Phase 2 (EOP2) meeting on 4/21/2005.

score in the 28 days prior to the final clinic visit were based on an average daily score, which was multiplied by 28 to obtain a value for the endpoint. However, if less than 14 days of data were available, the endpoint was considered missing and last observation carried forward was used to impute the outcome.

The statistical analyses of the co-primary endpoints did not necessarily include eDiary data that corresponded to Week 24. The eDiary data were 'anchored to clinic visits', even though the eDiary data were collected and transferred daily. As would be expected, clinic visits did not necessarily occur exactly on time or include a window of ±7 days. So, for example, if a subject came to clinic on Week 23 instead of Week 24, that visit was identified as Week 24 for purposes of data analyses.

The daily average of the number of SSEs entered into the eDiary in the 28 days prior to the clinic visit (without spanning into the previous clinic visit) was multiplied by 28 to arrive at the monthly (28 day) total count as shown in the following algorithm:

> Total monthly count of SSEs = 28 x [(sum of the number of SSEs entered) / (sum of number of days entered)]

For example, if a subject had entered data on 24 of the days since the last clinic visit, and had counted 6 SSEs, her monthly SSE score would be 28 x 6/24 = 7.

Similarly, the monthly sexual desire was calculated as:

> Monthly sexual desire = 28 x [(sum of desire scores) / (sum of number of days entered)]

2.1.3 Major design and statistical issues

In this section, I discuss issues with the design of the trials, analyses and results that are common to Studies 511.71 and 511.75, two of the Phase 3 studies that the Applicant is using to support the efficacy of flibanserin 100 mg q.h.s. These issues are critical to understanding and interpreting the results of the studies.

2.1.3.1 Statistically significant finding for only one of two co-primary endpoints

Because only one of the two co-primary endpoints demonstrated statistical significance in Studies 511.71 and 511.75, neither study is a positive study that supports the efficacy of flibanserin for the treatment of HSDD. This conclusion is consistent with the individual study protocols, the statistical analysis plans (SAPs), and the discussion between FDA and the Applicant at the End of Phase 2 meeting held on 4/21/2005. The protocols, SAPs, and meeting minutes indicate in order for a clinical study to have a successful outcome, a study needed to demonstrate both statistical and clinical significance for treatment versus placebo for each of the

co-primary endpoints.

Only one co-primary endpoint, satisfactory sexual events (SSE), demonstrated statistical significance for the comparisons between the flibanserin 100 mg q.h.s. and placebo treatment groups in Studies 511.71 and 511.75. The other co-primary endpoint, sexual desire (as recorded in the eDiary) was not statistically significant. Although both endpoints were statistically significant in Study 551.74, which was a randomized withdrawal study, I consider this study only supportive in establishing the efficacy of 100 mg q.h.s; see Section 2.1.1 for further details.

From a statistical perspective, therefore, the evidence in this NDA does not support the efficacy of flibanserin for the treatment of HSDD.

2.1.3.2 Study Discontinuations and Imputation for Missing Data

Given the modest treatment effects for SSE and the non-significant treatment effects for desire, the interpretation of the results of the studies must take into account the rates of study discontinuations and data imputation in these studies. The rate of study discontinuations was higher among subjects treated with flibanserin 100 mg q.h.s. than among subjects treated with placebo; 31% - flibanserin 100 mg q.h.s. vs 25% - placebo in both studies combined. This difference between the treatment groups in discontinuations resulted in a higher rate of data imputation among flibanserin-treated subjects.

Unfortunately, some of these discontinuations were an artifact of the study design. The protocols stipulated that subjects were to be discontinued from the study for:

- Any concomitant illness that prevents compliance, or
- Failure to take any study medication for more than seven consecutive days

Resulting directly from these provisions, study discontinuation rates and the rates of imputation for missing data are higher than needed. The number of discontinuations and the amount of data imputation could have been minimized by allowing subjects to remain in the studies regardless of the subjects' treatment status and compliance with study medication. In addition, some of these discontinuations likely represent study treatment failures due to lack of efficacy or intolerability.

Although the Applicant presents analyses suggesting the effect of study discontinuations on the study results was minimal, the underlying assumption is untestable. The analyses assume that those who dropped out at Week 8, for example, would have had the same result had they completed the 24-week study. The difference between the treatment groups in the rates of study discontinuations suggests, however, this assumption may not hold. Moreover, these provisions weaken the analyses that adopt the intent-to-treat principle.

2.1.3.3 Noncompliance on eDiary desire

The Applicant asserts the lack of validation for the eDiary measure of desire and the frequency of noncompliance for this item may partially explain the non-significant finding for this co-primary endpoint. The Applicant states the eDiary measure of desire was "introduced without extensive validation" and claims, *post hoc*, the sexual desire domain of the FSFI instrument is a more appropriate endpoint for assessing sexual desire.

The Applicant's briefing document notes "as the clinical program progressed, important limitations of the eDiary became evident. These included a) Sub-optimal compliance, b) limited response scale that effectively rendered it only a 3-level measure, and c) reactivity of the measure due to annoyance of daily completion[2]". The briefing document also points to a 24 hour recall period for eDiary sexual desire compared with a 72 hour recall period for SSE as a reason why noncompliance for the eDiary desire item was greater than noncompliance for SSE.

Noncompliance for the eDiary sexual desire endpoint, however, actually provides information on a subject's level of desire; noncompliance should not be treated as missing data. Analyses show subjects who did not respond to the eDiary desire item had a mean number of SSEs that fell between the mean SSEs for subjects reporting no desire and those reporting minimal desire. This pattern, which was seen at Weeks 4, 8, 12, 16, 20 and 24 in each of the two studies (see Section 3.1.1.3), suggests that a missing value on the eDiary desire item likely represents subjects whose level of desire falls between no desire and minimal desire.

2.1.3.4 Applicant's proposal to elevate FSFI-desire, a pre-specified secondary endpoint

The Applicant would like to use the results from FSFI-desire index in place of the eDiary desire item to support the efficacy of flibanserin for the outcome of sexual desire. If the FSFI-desire is substituted for eDiary desire as a co-primary endpoint, the results for desire become statistically significant for both studies. To support the change in endpoint, the submission states the Female Sexual Function Index – Desire Items (FSFI-desire) is a better measure than the eDiary desire item because FSFI-desire captures both frequency and intensity, and covers a longer recall period.

According to the submission, research conducted after the introduction of the eDiary shows the eDiary measurement is not concordant with a woman's own report of her experience with HSDD that, the submission states, includes both infrequency and lack of intensity of sexual desire over periods of time much in excess of a daily reporting time frame. The submission also indicates FSFI-desire "showed a highly consistent, although nominally significant ($p<0.01$), clinically meaningful impact of flibanserin treatment versus placebo" and "this finding demonstrates that flibanserin is more consistently effective in increasing women's global experience of desire than in increasing the intensity of their acute episodes of desire" as measured by the eDiary.[3] The

[2] Applicant's briefing document, dated 6/18/2010, page 32.
[3] Applicant's briefing document, dated 6/18/2010, page 20.

submission also states "taken together, this body of evidence provides consistent empirical and conceptual support for the FSFI assessment of sexual desire as a relevant and appropriate endpoint on sexual desire in clinical trials of women with HSDD[4]".

However, from a statistical perspective, this argument is problematic. First, FSFI-desire was not the prespecified co-primary endpoint. Although there can be situations where an endpoint other than the pre-specified endpoint could be used to assess the efficacy of a product, the substitution of a patient-reported outcome as a primary endpoint after the data are unblinded does not seem to be such a situation. Furthermore, the FSFI was part of a battery of tests that were administered during clinic visits. Subjects first completed the 13-item Female Sexual Distress Scale-Revised (FSDS-R) and then completed the FSFI. The submission does not address the effect that completing the FSDS-R may have had on a subject's responses to the FSFI.

2.1.3.5 Responder analyses and clinically meaningful changes

To help understand whether the changes observed for satisfactory sexual events and sexual desire were meaningful to the subjects enrolled in the studies, the Applicant conducted responder analyses. A responder analysis of patient-reported outcomes is a descriptive technique that is used to help inform whether changes observed during a clinical study are clinically meaningful or not. In a responder analysis, a minimum change from baseline, or threshold, defines whether a subject has experienced a clinically meaningful improvement or has not. This threshold dichotomizes the data, and a subject is classified as a responder or a non-responder.

The process for selecting the threshold plays a crucial role in responder analyses. However, responder analyses focus only on "successes". They do not consider subjects whose status deteriorated during a clinical trial. For example, subjects assigned to active treatment could experience adverse events and, thus, deteriorate from baseline at a rate greater than observed for subjects assigned to placebo. A responder analysis would not capture this scenario.

The medical division conveyed comments on responder analyses through a Special Protocol Assessment (SPA), dated 4/14/2006, for Study 511.70. Although the Applicant did not request SPAs for the other studies, the comments on responder analyses and clinically meaningful differences contained in the SPA for Study 511.70 apply to the other studies and form the basis for the analyses reported in the NDA.

According to the SPA, the medical division required a responder analysis for each of the co-primary endpoints. The SPA stated that an increase from baseline of two satisfactory events and four "desire days" per month appeared reasonable for defining a response. The SPA also indicated that a failure to demonstrate a statistically significant benefit of therapy over that of placebo in the responder analyses would be a review issue.

The SPA comments also requested the Applicant to submit to the Division for review the

[4] Applicant's briefing document, dated 6/18/2010, page 90.

procedure(s) by which the applicant would determine the degree of change that subjects perceive as being clinically meaningful, perhaps by including this assessment in at least one of the Phase 3 trials. Failure to provide evidence of the clinical meaningfulness of the change in desire would be a review issue. The Division also requested that the NDA include data to support the clinical meaningfulness of the increase in the number of SSEs in the active treatment groups compared to that in the placebo group, assuming that the primary analysis shows that the change is statistically significant.

The methods used to define "response" are discussed below in 3.1.1.3.

2.1.3.6 Visit windows defined for the analyses of the co-primary endpoints

Although the protocols called for analyses of change between the 4-week baseline and the four week window of Weeks 21 to 24, the Applicant's analyses of the co-primary endpoints did not adhere to this plan. Instead, the eDiary data were divided into 4-week intervals that were anchored to a subject's clinic visits, and the eDiary data for the co-primary endpoints were computed as a 28-day average of all data collected between visits.

As expected, clinic visits did not necessarily occur every four weeks; visits included a window of ±7 days. So, for example, if a subject came to clinic on Week 23 instead of Week 24, that visit was identified as Week 24 for purposes of data analyses. In that case, the data collected during the 28 days ending with Week 23 would have been designated as the eDiary data for Week 24. Moreover, if the first week of a 4-week interval overlapped with the 4-week interval of the previous clinic visit, then only 3-weeks of data were used for the 4-week interval.

Although this approach to creating windows is common to studies in which certain assessments are made during a clinic visit, the approach is not necessary when diary data are collected daily at home. In the example above, whereas the analyses used the 28 days ending with Week 23 in order to 'anchor' the diary data to the date of the clinic visit, the use of the 28 days ending with Week 24 would have been a more appropriate analysis of the co-primary endpoints.

Anchoring the eDiary data to clinic visits also has implications on the computation of the 28-day averages and the handling of missing data. The statistical analysis plans for Study 511.71 and Study 511.75 state[5]:

> "Patients may enter data for differing number of days since the protocol allows for 28±7 days. Therefore, the daily average of the counts (for each question) entered into the eDiary in the 28 days prior to the clinic visit (without spanning into the previous clinic visit) will be multiplied by 28 to arrive at the monthly total count."

According to the minutes of the pre-NDA meeting held on October 10, 2007, FDA anticipated

[5] Study 511.71 and Study 511.75: Clinical Study Report, Section 16.9.1 Statistical analysis plan and further statistical considerations, Section 5.1.1 Satisfying sexual event

this issue of varying number of days of eDiary data corresponding to a clinic visit. The FDA advised the Applicant to consider using data from the most proximal 28-day interval, regardless of its temporal relationship to a study visit. "The Division would encourage the Sponsor to consider using this sliding window for any 4-week treatment intervals that have <21 days of data entered in the diary. The objective of this recommendation was to carry forward the most current data and to avoid extensive extrapolation of results based on only 14-20 days of data.

Results of analyses conducted on 28-day intervals, starting with the time of randomization, were not as convincing as analyses conducted on the intervals that were anchored to study visits. As discussed in 3.1.1.3, the treatment effect in Study 511.75 was smaller with a p-value of 0.048. When all available data were used, regardless of the 14-day requirement, the treatment effect in Study 511.75 was statistically non-significant.

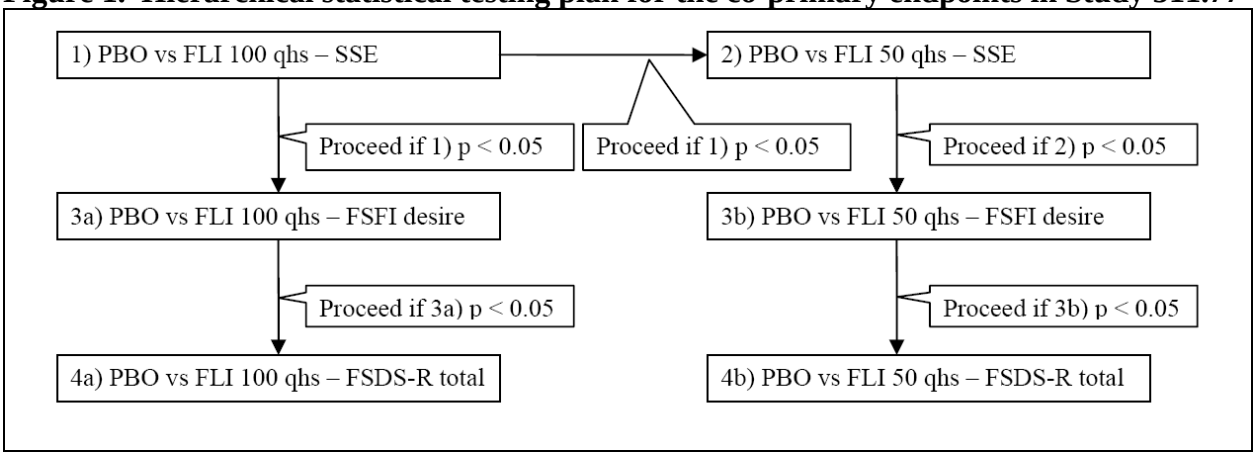
2.1.3.7 Overview of statistical methods used to control for multiple comparisons

The statistical analysis plans stipulated that both co-primary endpoints needed to be statistically significant in order for a trial to be considered a positive study. Pairwise comparisons using the Wilcoxon rank sum tests evaluated the satisfying sexual event endpoint and analysis of covariance evaluated the monthly sum or responses for sexual desire.

The statistical analysis plan for the North American Studies (Study 511.71, Study 511.75 and 511.74) used Hochberg's method to control for multiple comparisons. Hochberg's method was implemented in two stages; first for SSE and then for sexual desire. At the first stage, the procedure identified the doses that were significantly different from placebo for the endpoint of satisfying sexual encounters. Second, from among those doses that were statistically significant for SSEs, Hochberg's method identified the doses that were statistically significant for sexual desire.

The European study (Study 511.77) used a hierarchical approach instead of the Hochberg method; see Figure 1. For the SSE endpoint, flibanserin 100 q.h.s. versus placebo was tested first. If that result was statistically significant at 0.05, then flibanserin 50 q.h.s. versus placebo would be tested at 0.05. If the SSE was statistically significant for a dose, then FSFI desire would be assessed at 0.05. If that was statistically significant, then FSDS-R total would be assessed at 0.05.

Figure 1. Hierarchical statistical testing plan for the co-primary endpoints in Study 511.77



Source: Figure 7.5.1, Statistical Analysis Plan for Study 511.77

2.1.4 Summary of results

Using the pre-specified approaches for adjusting for multiplicity, outlined in 2.1.3.7, only Study 511.74 – the randomized, withdrawal study – was a positive study; see Table 2. None of the other studies won on both co-primary endpoints. Studies 511.71 and 511.75 were statistically significant for SSE but not sexual desire. Study 511.77 did not yield a statistically significant result for SSE; therefore, testing of other endpoints was not appropriate.

Table 2. Summary of results for the co-primary endpoints; flibanserin 100 mg q.h.s. and placebo treatment groups.

	Satisfying Sexual Events				Sexual Desire****			
	Median change from baseline				*Mean** change from baseline			
	Placebo	Flibanserin 100 mg q h.s.	Treatment Difference	p-value**	Placebo	Flibanserin 100 mg q.h.s.	Treatment Difference	p-value***
Study 511.71	0.0	1.0	1.0	<0.05	6.9	9.1	2.2	>0.05
Study 511.75	0.5	1.0	0.5	<0.05	6.8	8.5	1.7	>0.05
Study 511.74	-2.3	-1.4	-1.0	<0.05	29.2	25.4	-3.8	<0.05
Study 511.77	0.0	1.0	1.0	>0.05	0.5	0.7	0.1	Not applicable, as specified by hierarchical testing procedure, because SSE was non-significant

***** Least-squared means, estimated from analysis of covariance models
****** Wilcoxon rank sum test, except for Study 511.74. ANCOVA used for 511.74; LS means are reported
*** Analysis of covariance, adjusting for center and baseline, except for Study 511.74. Longitudinal models used for 511.74; LS means are reported.
**** eDiary Sexual Desire was the endpoint in Studies 511.71, 511.75 and 511.74; FSFI-desire was the endpoint In Study 511.77.
Note: Because Hochberg's method was used to adjust for multiple treatment comparisons, the p-values are reported as >0.05 or <0.05.

Source: Tables 15.1.1:2 and 15.2.1:5, Clinical Study Report for 511.71; Tables 15.2.1:2 and 15.2.1:5, Clinical Study Report for 511.75; Tables 15.2.1.1:4 and 15.2.1.2:3, Clinical Study Report for 511.74; Tables 15.2.1:2 and 15.2.1:9, Clinical Study Report for 511.77.

2.2 Data Sources

- Electronic submission: \\CDSESUB1\EVSPROD\NDA022526
- Datasets
- Special Protocol Assessment (Study 511.70) \\cdsesub1\evsprod\NDA022526\0000\m1\us\2006-04-14-ind-(b) (4)-spa-511-70.pdf
- Response to Information Request (Response dated 4/30/2010, Sequence 0018) \\cdsesub1\EVSPROD\NDA022526\0018\m1\us\signed-cover-letter-0018.pdf
- Response to Information Request (Response dated 5/12/2010, Sequence 0019) – eDiary compliance information and discontinuations by study visit \\cdsesub1\EVSPROD\NDA022526\0019\m1\us\signed-cover-letter-0019.pdf
- Response to Information Request (Response dated 5/7/2010, Sequence 0020) \\cdsesub1\EVSPROD\NDA022526\0020\m1\us\signed-cover-letter-0020.pdf

- Response to Information Request (Response dated 5/12/2010, Sequence 0021) – how to (1) calculate days between randomization and clinic visits and (2) calculate days on study treatment \\cdsesub1\EVSPROD\NDA022526\0021\m1\us\signed-cover-letter-0021.pdf
- Response to Information Request (Response dated 5/14/2010, Sequence 0022) – description of criteria and methods for discontinuing subjects, anchoring eDiary data to clinic visits, re-analyses using blocks of 28 days starting at time of randomization, analyses of the relationship between "not reported" on 'eDiary desire' and SSEs \\cdsesub1\EVSPROD\NDA022526\0022\m1\us\signed-cover-letter-0022.pdf
- Response to Information Request (Response dated 6/16/2010, Sequence 0029) – clarification on analyses of the relationship between "not reported" on 'eDiary desire' and SSEs \\cdsesub1\EVSPROD\NDA022526\0029\m1\us\signed-cover-letter-0029.pdf
- BI's method for determining the responder criteria definition \\Cdsesub1\evsprod\NDA022526\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\hsdd-women\5353-rep-analys-data-more-one-stud\rcm
- Applicant's Background Package for the Advisory Committee meeting, 6/18/2010 http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ReproductiveHealthDrugsAdvisoryCommittee/UCM215438.pdf

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Phase 3 studies: 511.71 and 511.75

3.1.1.1 Study Design

Both studies were 24-week, randomized, double-blind, placebo-controlled studies of flibanserin in premenopausal women with hypoactive sexual desire disorder.

In Study 511.71, subjects were randomized to one of three treatment groups:
- Placebo
- Flibanserin 50 mg q.h.s.
- Flibanserin 100 mg q.h.s.

In Study 511.75, subjects were randomized to one of four treatment groups:
- Placebo
- Flibanserin 25 mg BID
- Flibanserin 50 mg q.h.s. for two weeks, then titrated to Flibanserin 50 mg BID
- Flibanserin 50 mg q.h.s. for two weeks, then titrated to Flibanserin 100 mg q.h.s

The entry criteria for Study 511.71, Study 511.74, Study 511.75 and Study 511.77 included:
- Premenopausal women

- 18 years of age and older
- Primary diagnosis of HSDD, generalized acquired type, according to DSM-IV-TR criteria. The current episode must be at least 24 weeks in duration by the baseline visit.
- A score of "15" or higher on the FSDS-R at the Screen Visit

An additional criterion for Study 511.71 was:

- Item Number Two of the Sexual Interest and Desire Inventory – Female (SIDI-F) must be rated as "0" or "1" at the Screen Visit

These studies excluded women with other diagnoses of sexual disorder or dysfunction, major depressive disorder or disorders that could affect sexual function, and women who started psychotherapeutic (non-drug) treatment within 12 weeks of the baseline visit.

3.1.1.2 Description of Subjects

Enrollment and Demographics

A total of 2,462 premenopausal women were randomized to treatment with placebo or with flibanserin at different doses in Studies 511.71 and 511.75; see Table 3. Study 511.70 randomized 1392 subjects.

Table 3. Number of subjects randomized, by treatment and study

Treatment	Study		
	Study 511.71	Study 511.75	Study 511.70
Placebo	295	399	350
25 mg bid	-	396	340
50 mg q.h.s	295	-	365
50 mg bid	-	393	337
100 mg q.h.s	290	396	-
Total	880	1582	1392

The demographics of the subjects enrolled in Studies 511.70, 511.71 and 511.75 were comparable; see Table 4. The typical subject was 35 years old, white, married, in her present relationship for 10 years, and a non-smoker. Around 10% of subjects were 45 years or older, and around 10% were black.

Table 4. Demographics

		511.70		511.71		511.75		Pooled (511.71/.75)	
		Placebo	Flibanserin	Placebo	Flibanserin	Placebo	Flibanserin	Placebo	Flibanserin
Number of patients		349	1036	295	585	398	1183	693	1768
Age Group [N (%)]	18-34	158 (45.3)	468 (45.2)	126 (42.7)	261 (44.6)	157 (39.4)	531 (44.9)	283 (40.8)	792 (44.8)
	35-44	164 (47.0)	471 (45.5)	136 (46.1)	243 (41.5)	195 (49.0)	529 (44.7)	331 (47.8)	772 (43.7)
	>= 45 yrs	27 (7.7)	97 (9.4)	33 (11.2)	81 (13.8)	46 (11.6)	123 (10.4)	79 (11.4)	204 (11.5)
Age [years]	Mean	34.9	35.0	35.5	36.0	36.2	35.3	35.9	35.5
	SD	6.9	7.0	7.0	7.4	6.6	6.9	6.8	7.1
Race [N (%)]	White	297 (85.1)	912 (88.0)	256 (86.8)	512 (87.5)	370 (93.0)	1076 (91.0)	626 (90.3)	1588 (89.8)
	Black	49 (14.0)	105 (10.1)	33 (11.2)	66 (11.3)	22 (5.5)	90 (7.6)	55 (7.9)	156 (8.8)
	Asian	3 (0.9)	19 (1.8)	6 (2.0)	7 (1.2)	6 (1.5)	17 (1.4)	12 (1.7)	24 (1.4)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Ethnic Origin-Hispanic	No	328 (94.0)	969 (93.5)	272 (92.2)	536 (91.6)	370 (93.0)	1126 (95.2)	642 (92.6)	1662 (94.0)
	Yes	21 (6.0)	66 (6.4)	23 (7.8)	48 (8.2)	28 (7.0)	57 (4.8)	51 (7.4)	105 (5.9)
	Missing	0 (0.0)	1 (0.1)	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Current Marital Status	Unmarried	73 (20.9)	228 (22.0)	58 (19.7)	142 (24.3)	60 (15.1)	270 (22.8)	118 (17.0)	412 (23.3)
	Married	276 (79.1)	808 (78.0)	237 (80.3)	443 (75.7)	338 (84.9)	913 (77.2)	575 (83.0)	1356 (76.7)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
How long in present relationship (months)	Mean	9.9	10.5	10.4	10.8	11.4	10.5	11.0	10.6
	SD	6.2	6.6	6.5	6.7	6.2	6.5	6.3	6.5
Hormonal contraceptive use	HC non-users	205 (58.7)	583 (56.3)	181 (61.4)	346 (59.1)	239 (60.1)	726 (61.4)	420 (60.6)	1072 (60.6)
	HC users	144 (41.3)	453 (43.7)	114 (38.6)	239 (40.9)	159 (39.9)	457 (38.6)	273 (39.4)	969 (39.4)
Smoking history	Never smoked	234 (67.0)	713 (68.8)	200 (67.8)	408 (69.7)	258 (64.8)	769 (65.0)	458 (66.1)	1177 (66.6)
	Ex-smoker	63 (18.1)	197 (19.0)	60 (20.3)	111 (19.0)	91 (22.9)	266 (22.5)	151 (21.8)	377 (21.3)
	Current smoker	52 (14.9)	126 (12.2)	35 (11.9)	66 (11.3)	49 (12.3)	148 (12.5)	84 (12.1)	214 (12.1)

Source: Appendix 6, Table 3.1.2.1

Notes:

(1) 'Years' is the unit of time for 'How long in present relationship'; 'months' is a typographical error.
(2) The number of subjects in Study 511.70 excludes 7 subjects who were randomized but did not receive study drug

Source: Table 1.4:1, Summary of Clinical Efficacy

Distribution of SSEs and eDiary Desire at baseline

Across Study 511.71 and Study 511.75, the distributions of baseline SSEs and eDiary desire were similar; see Figure 2 and Figure 3. The distributions were skewed to the right. The number of SSEs reported at baseline ranged from 0 to 23, with a median count of 2 events. Approximately 20% of subjects reported no SSEs at baseline and around 25% of subjects reported at least 5 events at baseline.

The eDiary desire scores at baseline ranged from 0 to 54 with an average value of approximately 12 and a median of 9. Approximately 30% reported a baseline eDiary desire score of 5 or less and 20% reported a score of at least 25. The possible values for eDiary desire were 0 to 84.

Figure 2. Distribution of SSEs at baseline, by study

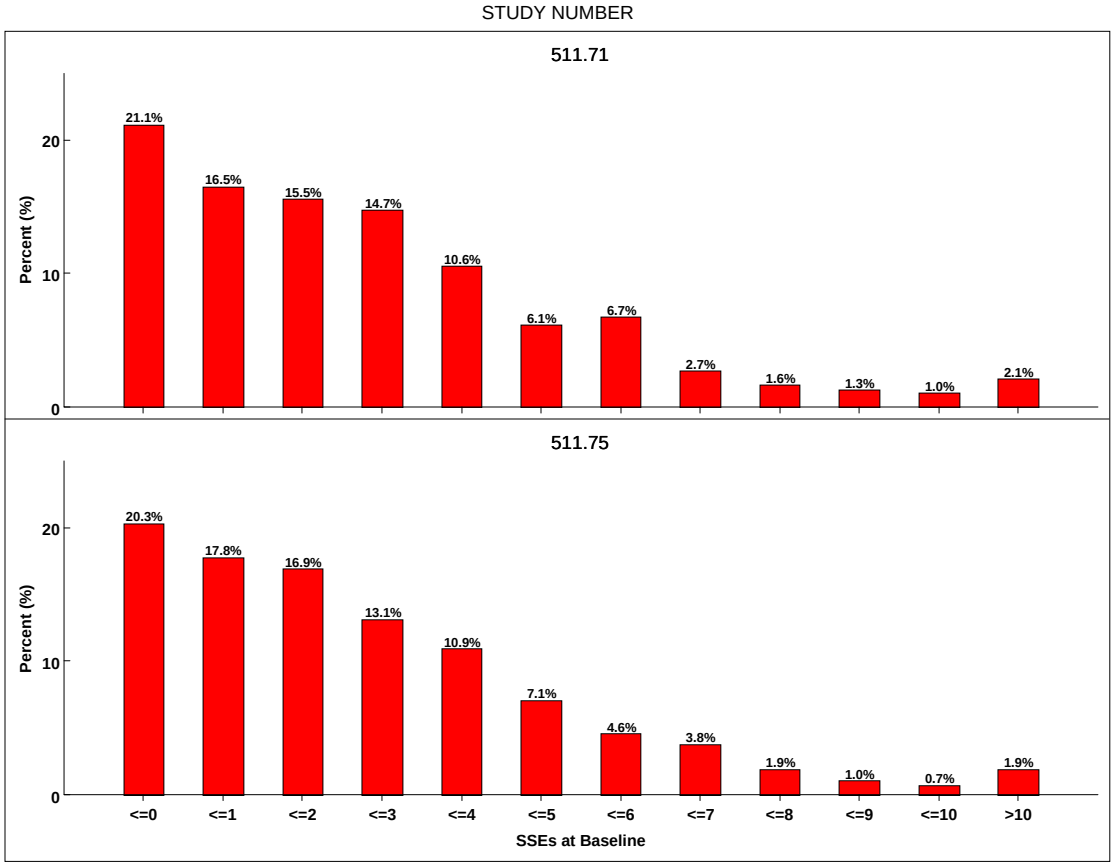
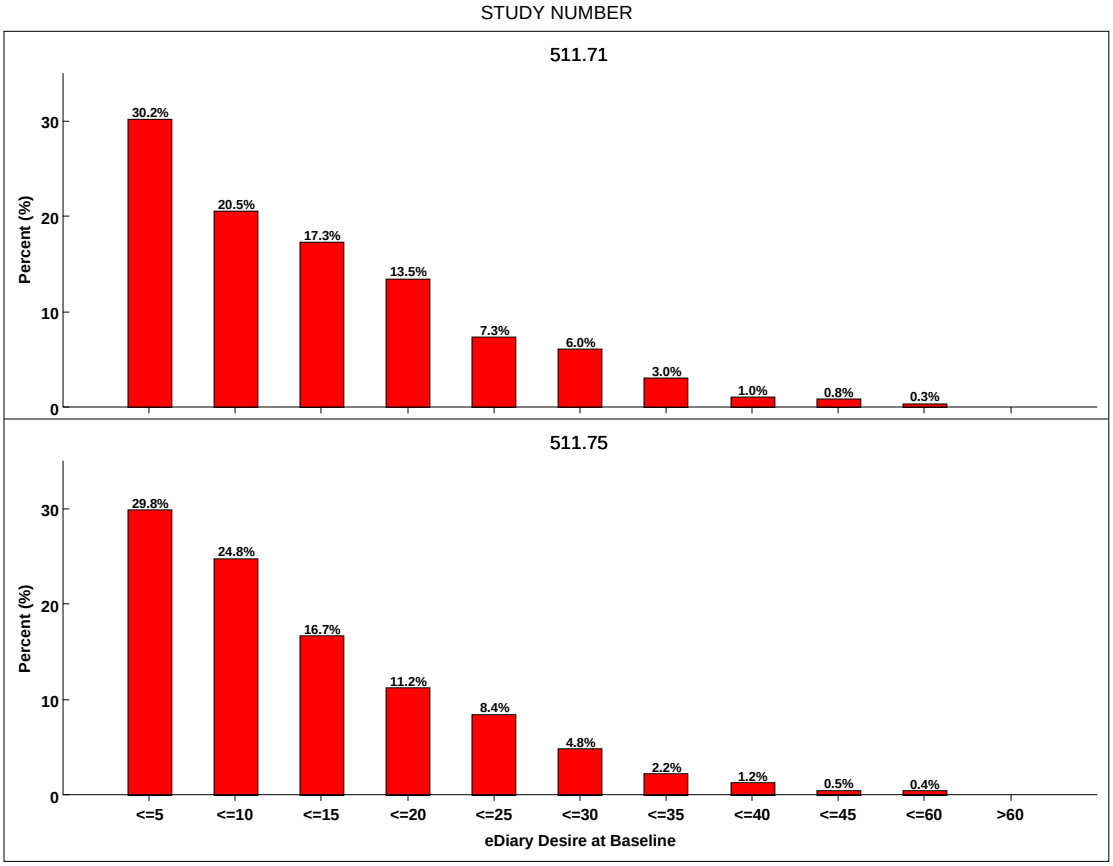


Figure 3. Distribution of eDiary desire at baseline, by study

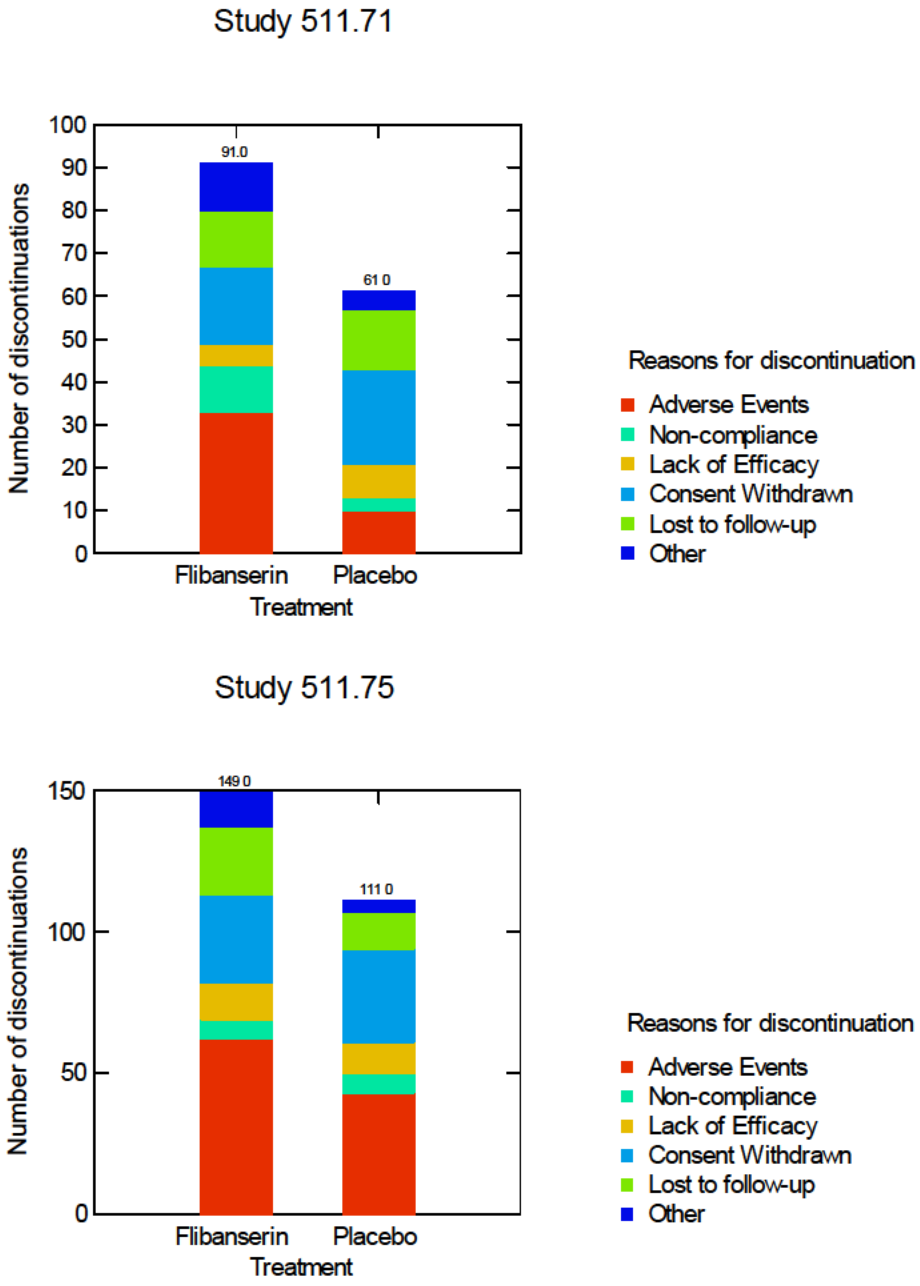


Disposition of Subjects

In each study, approximately 10% more study discontinuations occurred among subjects randomized to flibanserin 100 mg q.h.s. than among subjects randomized to placebo; see Figure 4. 'Adverse events' and 'Other' were the primary sources for this 10% difference in discontinuation rates. Additionally, in Study 511.71, a greater proportion discontinued due to 'Non-compliance', which was a protocol-specified reason for discontinuing from the study. Potentially, discontinuations for 'non-compliance' may have been related to the tolerability of flibanserin.

In Study 511.71, approximately 21% of placebo-treated subjects discontinued compared with 31% of subjects randomized to flibanserin 100 mg q.h.s. The overall discontinuations rates in Study 511.75 were somewhat higher. Approximately 28% of placebo-treated subjects discontinued compared with 37% of subjects randomized to flibanserin 100 mg q.h.s.

Figure 4. Reasons for discontinuation from flibanserin 100 mg q.h.s. and placebo; by study

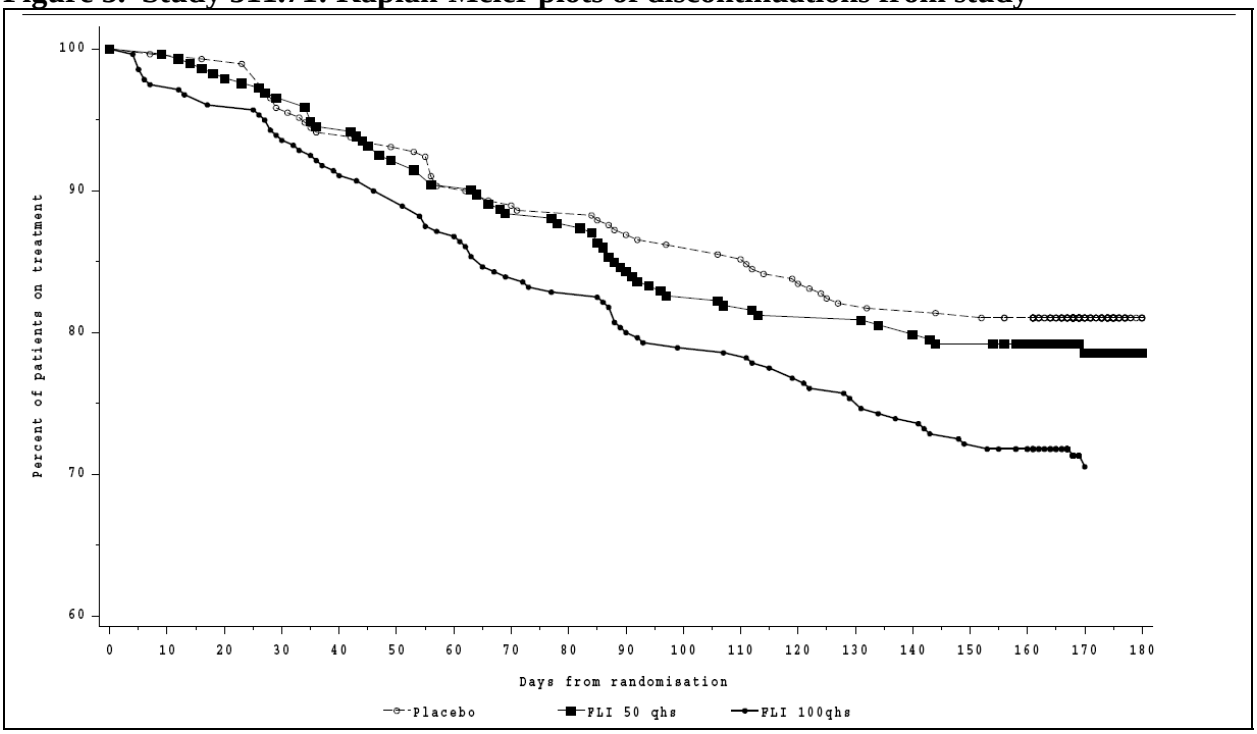


In Study 511.71, Kaplan-Meier graphs of study discontinuations, both overall and due to adverse events, show higher rates of discontinuations for flibanserin 100 mg q.h.s. than for flibanserin 50 mg q.h.s. and for placebo; see Figure 5 and Figure 6. Discontinuations due to adverse events

appear to be dose-related.

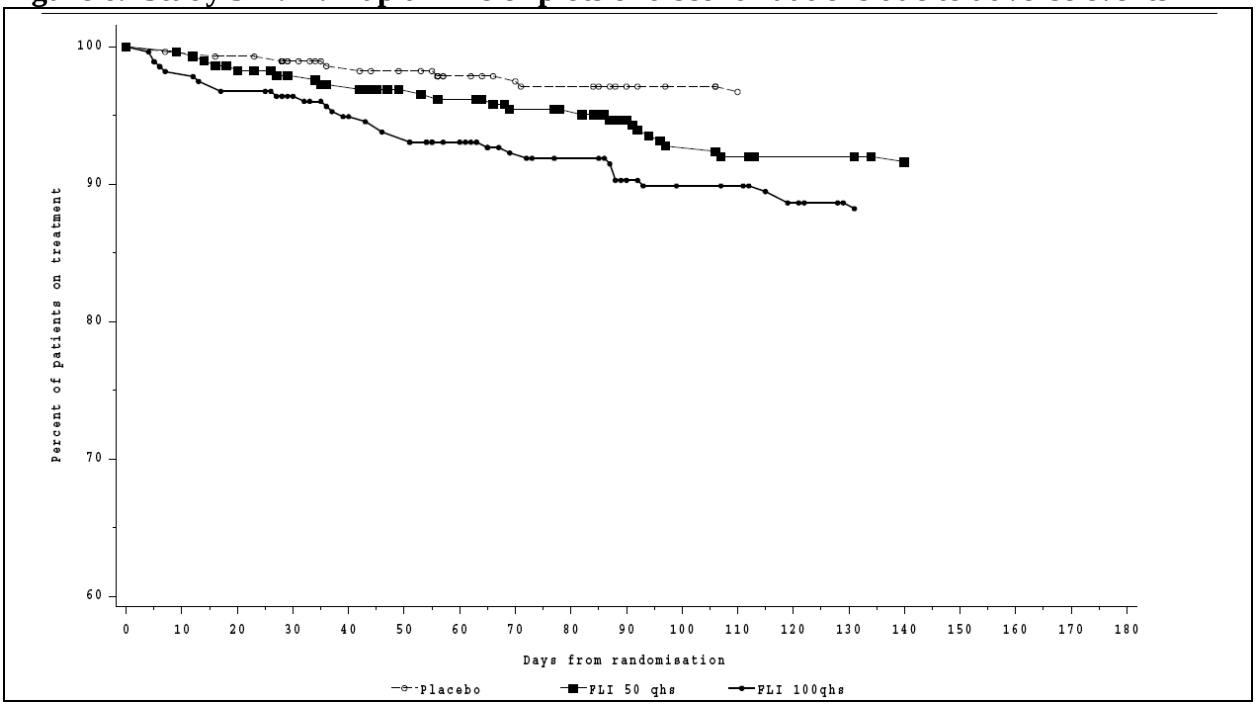
In Study 511.75, the differences among treatment groups are less pronounced. The graphs for flibanserin 50 mg bid and for flibanserin 100 mg q.h.s. are similar, with both showing higher rates of discontinuations when compared with the other treatment groups; see Figure 7 and Figure 8. Curiously, the rates of discontinuations due to adverse events are higher for flibanserin 50 mg bid than for flibanserin 100 mg q.h.s.

Figure 5. Study 511.71: Kaplan-Meier plots of discontinuations from study



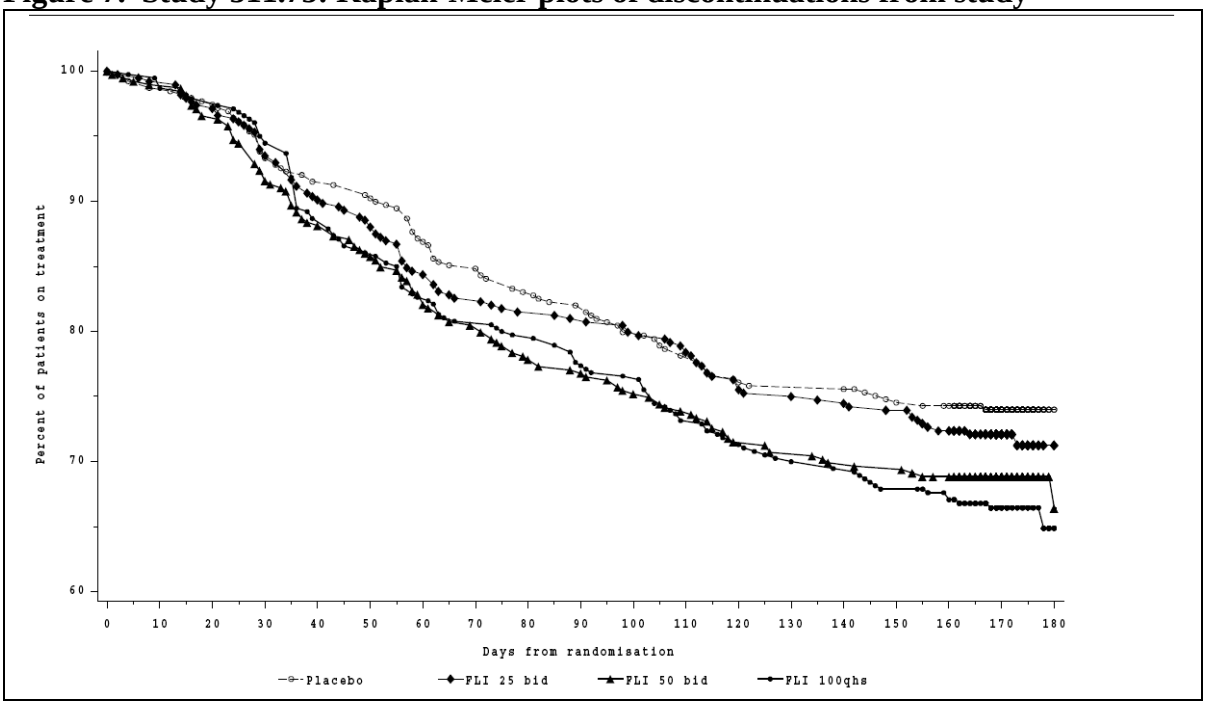
Source: Figure 15.1.1:1, Clinical Study Report for Study 511.71

Figure 6. Study 511.71: Kaplan-Meier plots of discontinuations due to adverse events



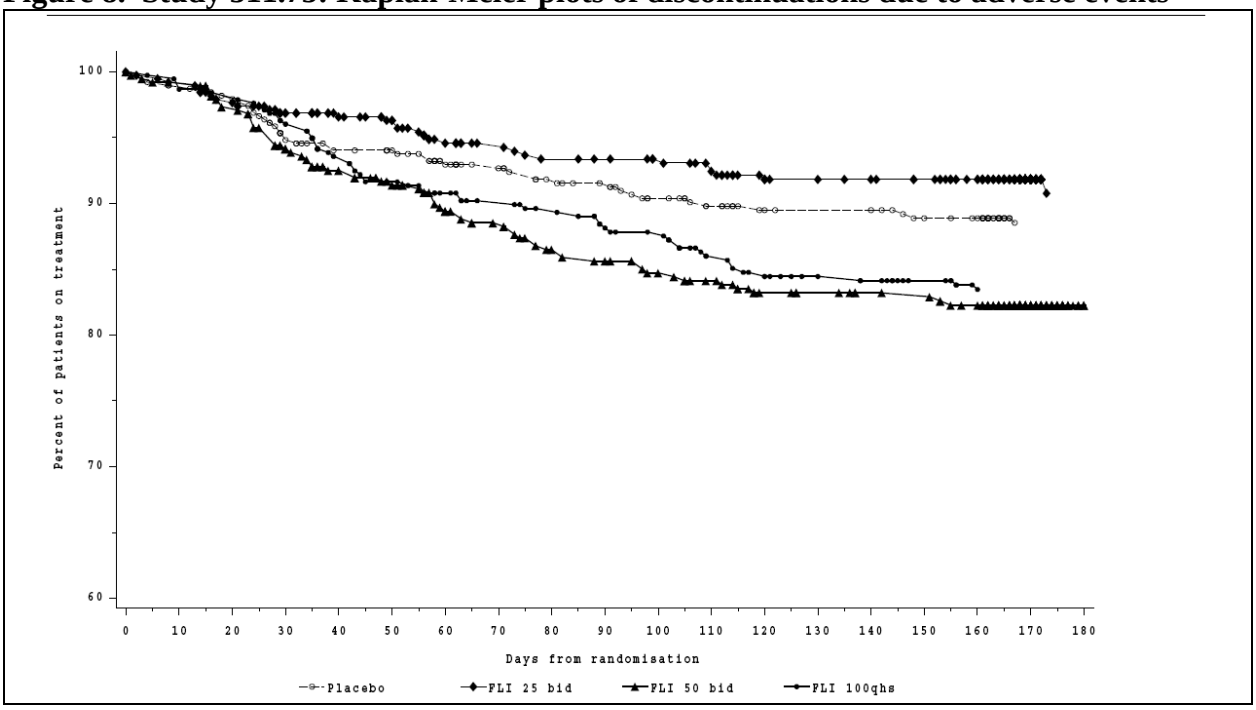
Source: Figure 15.1.1:2, Clinical Study Report for Study 511.71

Figure 7. Study 511.75: Kaplan-Meier plots of discontinuations from study



Source: Figure 15.1.1:1, Clinical Study Report for Study 511.75

Figure 8. Study 511.75: Kaplan-Meier plots of discontinuations due to adverse events



Source: Figure 15.1.1:2, Clinical Study Report for Study 511.75

3.1.1.3 Results

Efficacy

The results from the two Phase 3 studies were comparable; see Table 5. In each study, the treatment comparisons between flibanserin 100 mg q.h.s. and placebo were statistically significant ($p<0.05$), with a median treatment difference of 1.0 SSE for Study 511.7 and 0.5 SSE for Study 511.75. Although the NDA focuses on the means when reporting results for SSEs, I am reporting the medians because the analyses used the Wilcoxon rank sum test, which is a nonparametric equivalent to the t-test. The Wilcoxon rank sum test tests the null hypothesis that the observations in the two groups come from the same distribution versus a shift in the distributions.

The treatment difference of around two units for eDiary desire, the second of the two co-primary endpoints, was not statistically significant for either study ($p>0.05$, ANCOVA).

Table 5. Results for the co-primary endpoints, by study.

	Satisfying Sexual Events				Sexual Desire			
	Median change from baseline				LS Mean change from baseline*			
	Placebo	FLI 100 q h.s.	Treatment Difference	p-value**	Placebo	FLI 100 q h.s.	Treatment Difference	p-value***
Study 511.71	0.0	1.0	1.0	<0.05	6.9	9.1	2.2	>0.05
Study 511.75	0.5	1.0	0.5	<0.05	6.8	8.5	1.7	>0.05

***** Least-squared means, estimated from analysis of covariance models
****** Wilcoxon rank sum test
*** Analysis of covariance, adjusting for center and baseline
Note: Because Hochberg's method was used to adjust for multiple treatment comparisons, the p-values are reported as >0.05 or <0.05.

Source: Clinical Study Report for Study 511.71: Table 15.2.1:2, Table 15.2.1:5; Clinical Study Report for Study 511.75: Table 15.2.1:2, Table 15.2.1:5

The monthly SSE count and eDiary sexual desire score were based on the number of days with available data:

- Total monthly count of SSEs = 28 x (sum of the number of SSEs entered)/(sum of number of days entered)
- Monthly sexual desire = 28 x (sum of desire scores)/ (sum of number of days entered)

Consequently, the analyses considered a subject with 14 days of data to be the same as a subject with 28 days of data. In other words, the analyses did not take into account the number of days with missing data. To address this concern, the Applicant repeated the analyses using the unstandardized sum of the number of SSEs. (Similar analyses for the unstandardized sum of

desire scores were not provided.) The results for the total count of SSEs were consistent with those for SSEs standardized to 28 days; see Table 6.

Table 6. Results for total count of SSEs (not standardized over 28 days)

	Satisfying Sexual Events			
	Median change from baseline			
	Placebo	FLI 100 q h.s.	Treatment Difference	p-value*
Study 511.71	0.0	1.0	1.0	<0.05
Study 511.75	0.0	1.0	1.0	<0.05

***** Wilcoxon rank sum test

Note: Because Hochberg's method was used to adjust for multiple treatment comparisons, the p-values are reported as >0.05 or <0.05.

Source: Clinical Study Report for Study 511.71: Table 15.2.2.1.1:5; Clinical Study Report for Study 511.75: Table 15.2.2.1.1:5

The emphasis on point estimates of change from baseline belies the variability in responses as shown in Figure 9 and Figure 10. The change from baseline for SSEs ranged from a decrease of 17 events to an increase of 48 events. The change from baseline for eDiary Desire ranged from a decrease of 41 units to an increase of 84. For each endpoint, around 40% of subjects either remained the same or decreased from baseline. Note that a decrease could only occur among subjects who reported at least one SSE at baseline or a score of at least one for eDiary Desire.

Figure 9. SSE: Distribution of change from baseline, by study

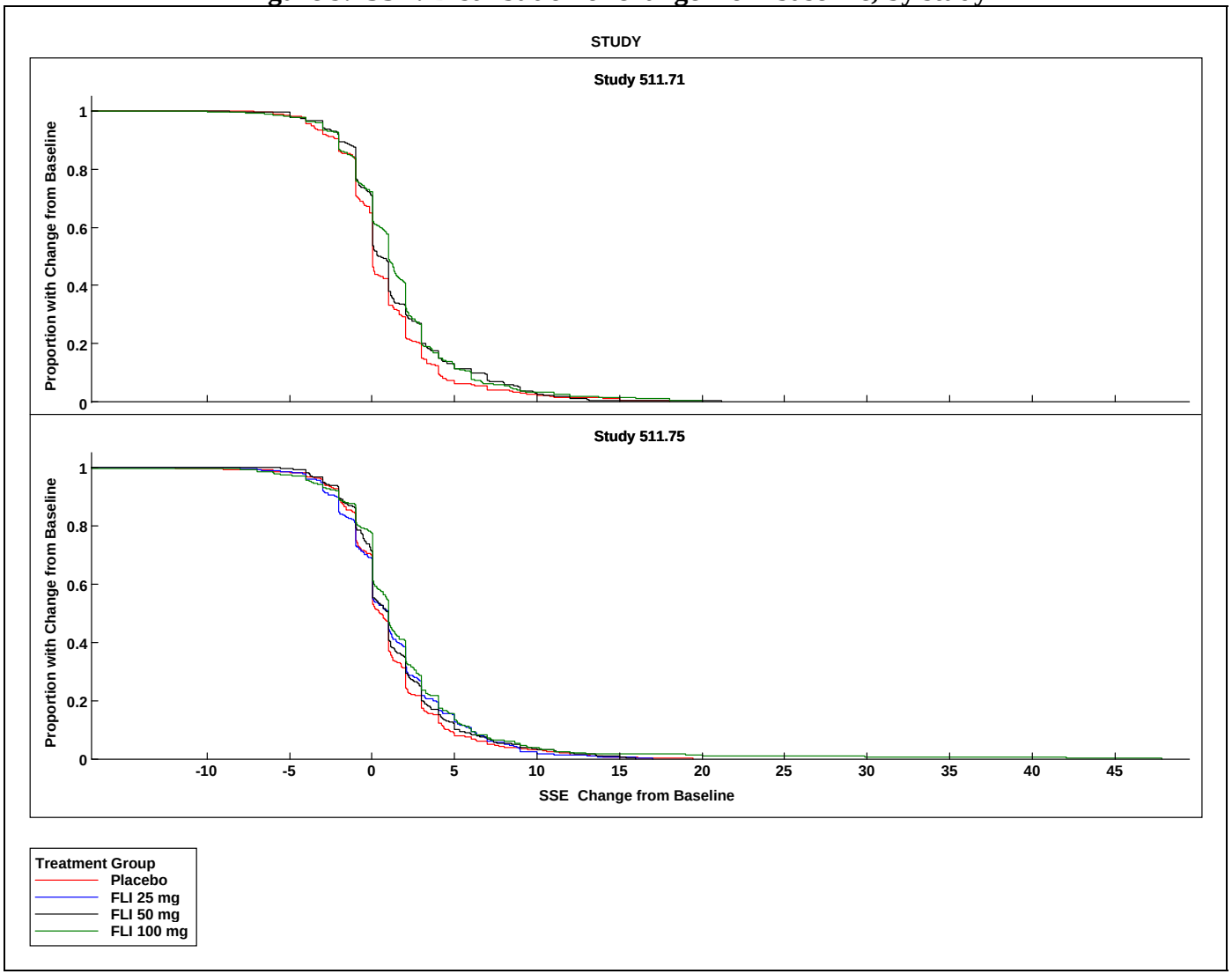
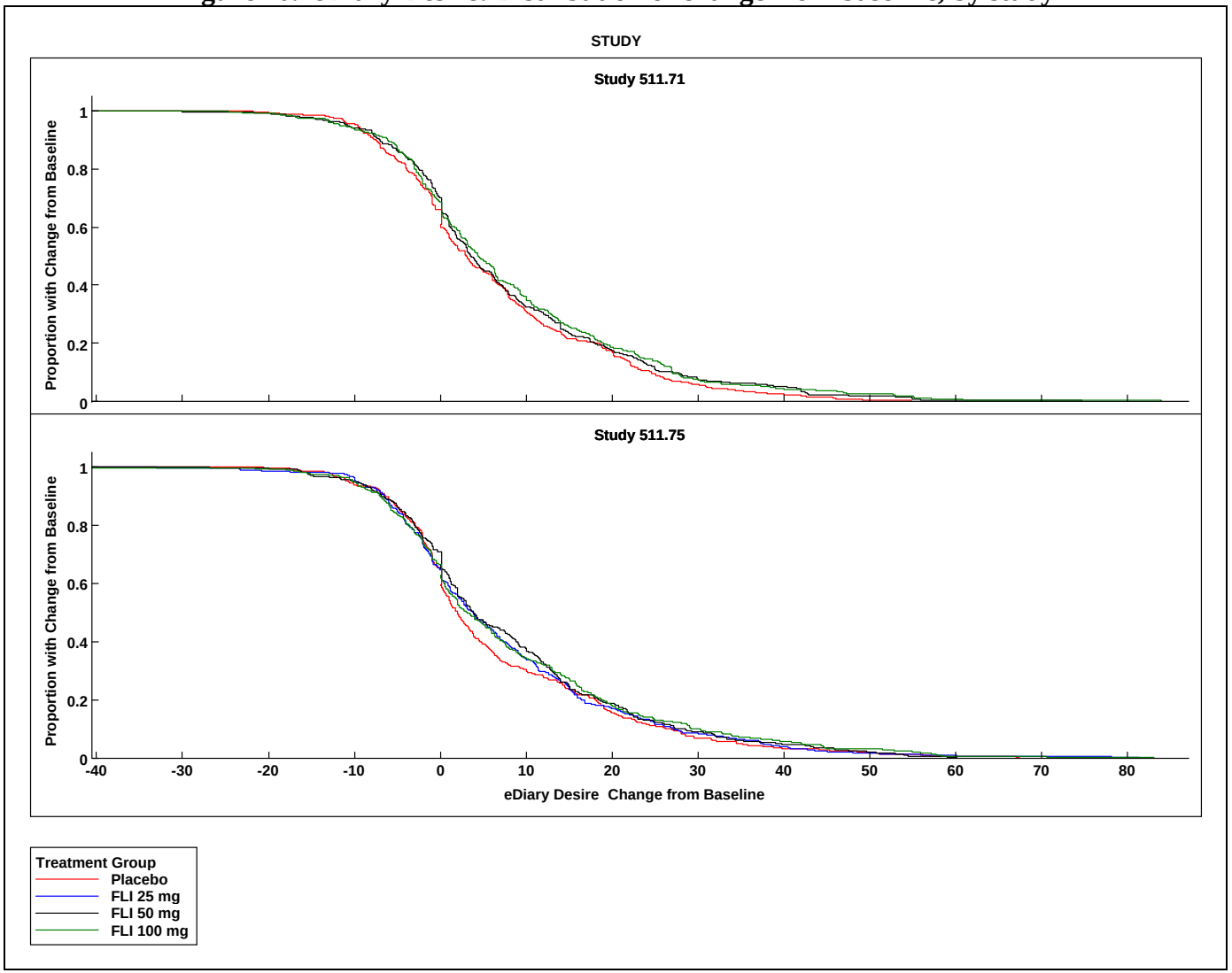


Figure 10. eDiary Desire: Distribution of change from baseline, by study



To explore whether change from baseline depended on subjects' baseline values, I did additional graphical analyses. First, I pooled all subjects (n=2393) from the two studies and calculated both the median and quartiles for SSE at baseline and eDiary Desire at baseline. Using these values, I redid Figure 9 and Figure 10 for subgroups defined by the median and quartiles. These graphs are limited to placebo and flibanserin 100 mg q.h.s.

The results for SSE do not appear to differ among the subgroups defined by the median (2 SSEs; see Figure 11) or among subgroups defined by quartiles (see Figure 14 in Appendix II).

The results for eDiary Desire are not as straightforward. Examination of the graphs for change from baseline suggest there may be an effect among subjects whose baseline eDiary Desire score was less than the median value of 9.33 while there may not be any effect among subjects whose eDiary Desire score at baseline was above the median; see Figure 12. This observation is more pronounced when the change from baseline is categorized by the quartile values of eDiary Desire at baseline; Figure 15. Among subjects in the first quartile (eDiary Desire score of 4 or below) those randomized to flibanserin 100 mg q.h.s. had a greater change than those randomized to placebo.

Figure 11. SSE: Distribution of change from baseline, by study and median SSE at baseline

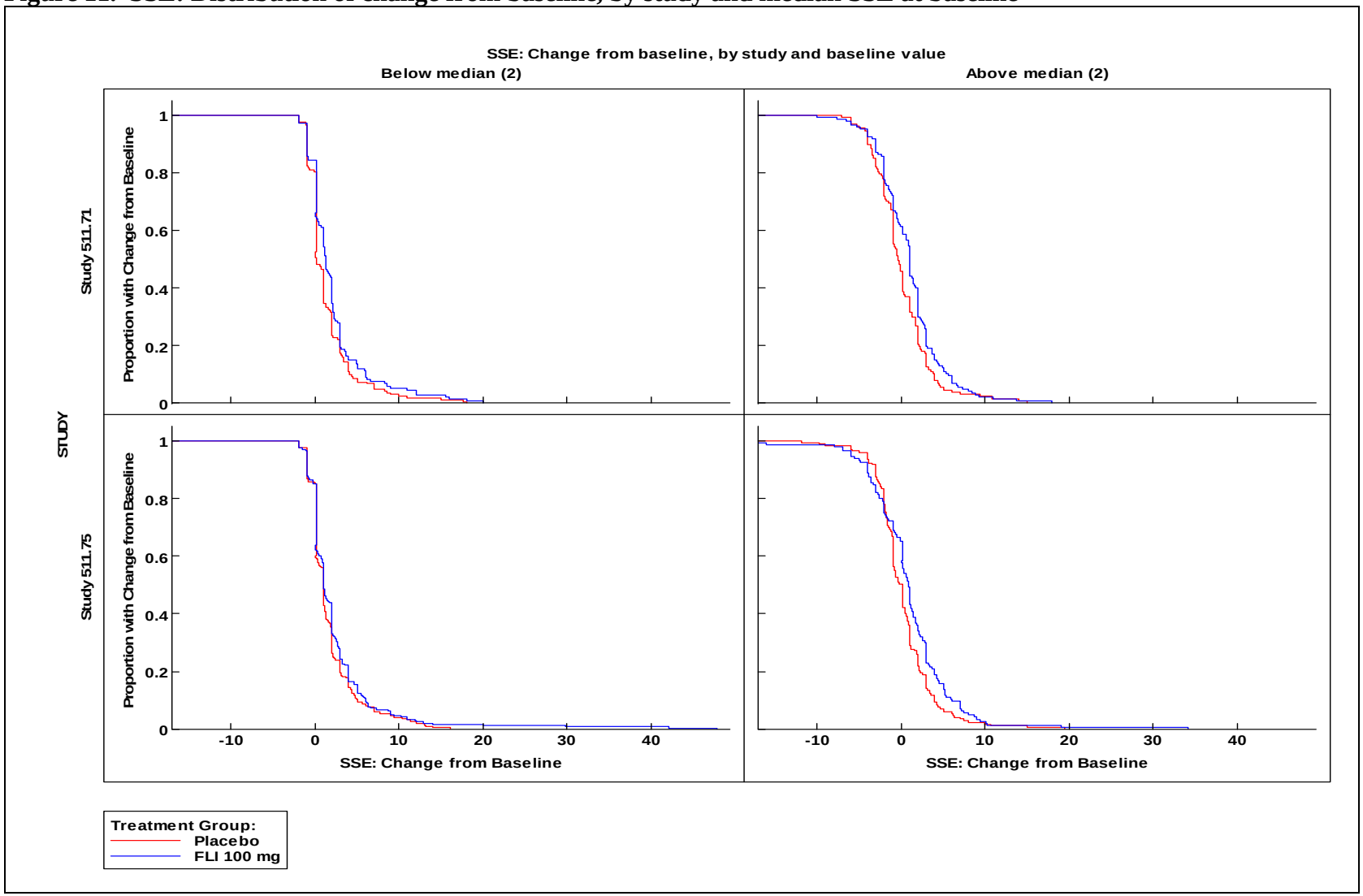
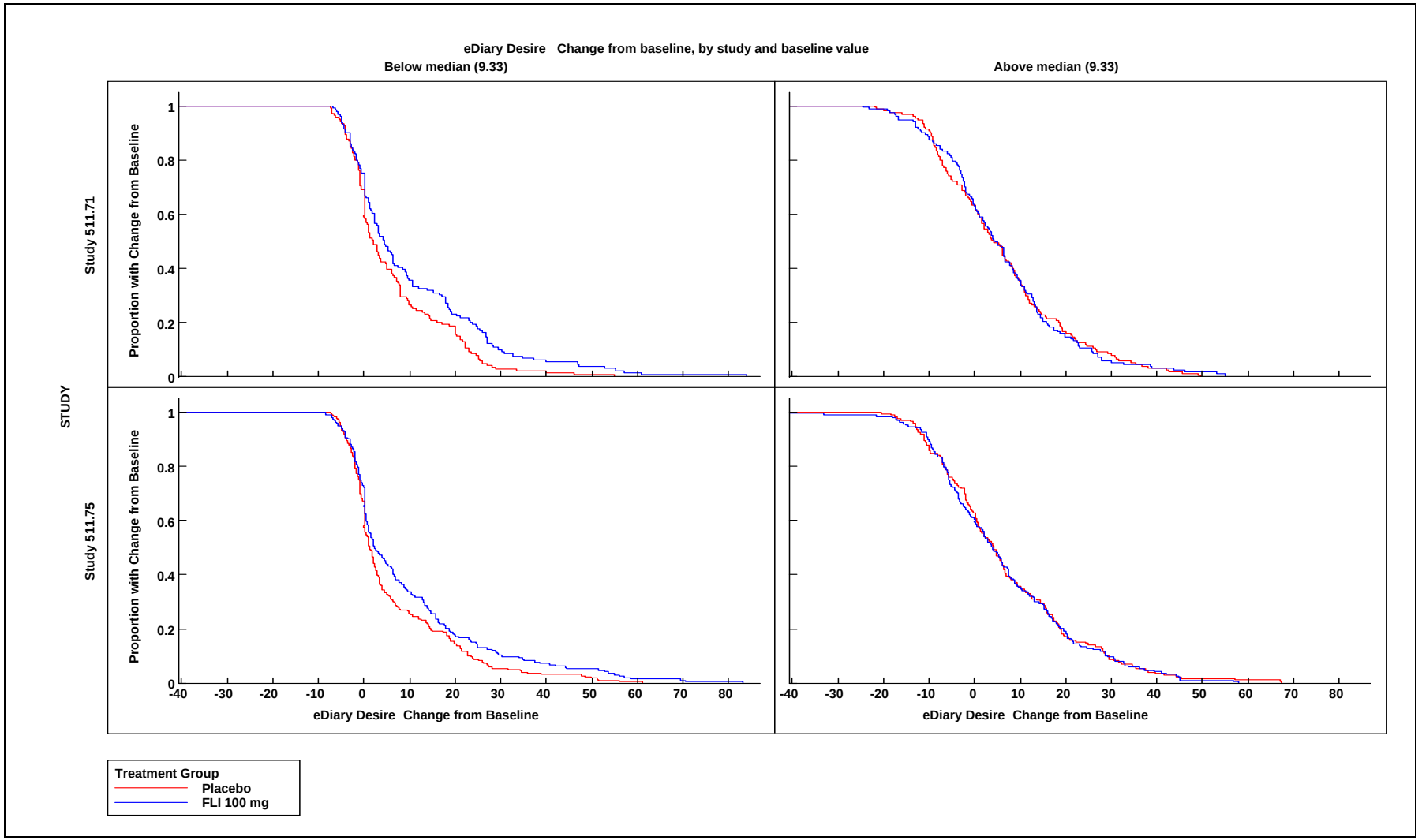


Figure 12. eDiary Desire: Distribution of change from baseline, by study and median eDiary Desire at baseline



Anchoring diary data to clinic visits

Although the protocols called for analyses of change between the 4-week baseline and Weeks 21 to 24, the Applicant's analyses of the co-primary endpoints did not adhere to this plan. Instead, the eDiary data were divided into 4-week intervals that were anchored to a subject's clinic visits, and the eDiary data for the co-primary endpoints were computed as a 28-day average of all data collected between visits.

As expected, clinic visits did not necessarily occur every four weeks; visits included a window of ±7 days. So, for example, if a subject came to clinic on Week 23 instead of Week 24, that visit was identified as Week 24 for purposes of data analyses. In that case, the data collected during the 28 days ending with Week 23 would have been designated as the eDiary data for Week 24. Moreover, if the first week of a 4-week interval overlapped with the 4-week interval of the previous clinic visit, then only 3-weeks of data were used for the 4-week interval.

Although this approach to creating windows is common to studies in which certain assessments are made during a clinic visit, the approach is not necessary when diary data are collected daily at home. In the example above, whereas the analyses used the 28 days ending with Week 23 in order to 'anchor' the diary data to the date of the clinic visit, the use of the 28 days ending with Week 24 would have been a more appropriate analysis of the co-primary endpoints.

Anchoring the eDiary data to clinic visits also has implications on the computation of the 28-day averages and the handling of missing data. The statistical analysis plans for Study 511.71 and Study 511.75 state[6]:

> "Patients may enter data for differing number of days since the protocol allows for 28±7 days. Therefore, the daily average of the counts (for each question) entered into the eDiary in the 28 days prior to the clinic visit (without spanning into the previous clinic visit) will be multiplied by 28 to arrive at the monthly total count."

According to the minutes of the pre-NDA meeting held on October 10, 2007, FDA anticipated this issue of varying number of days of eDiary data corresponding to a clinic visit. The FDA advised the Applicant to consider using data from the most proximal 28-day interval, regardless of its temporal relationship to a study visit. "The Division would encourage the Sponsor to consider using this sliding window for any 4-week treatment intervals that have <21 days of data entered in the diary. The objective of this recommendation was to carry forward the most current data and to avoid extensive extrapolation of results based on only 14-20 days of data.

We requested the Applicant to create 28-day intervals, starting with the day of randomization, that were not anchored to clinic visits. Using these intervals, the Applicant repeated the analyses of the co-primary endpoints, both for the endpoints standardized to 28 days (Table 7) and for the

[6] Study 511.71: Clinical Study Report, Section 16.9.1 Statistical analysis plan and further statistical considerations, Section 5.1.1 Satisfying sexual event

actual values observed (Table 8). Although the results for the data standardized to 28 days are consistent with the original analyses, the treatment difference for SSEs in Study 511.75 moves closer to statistical non-significance (p=0.048, adjusted for Hochberg). When all available data are used, the treatment difference is smaller than that observed when the results are restricted to 14-day minimum using LOCF and is non-significant.

Table 7. Results for the co-primary endpoints standardized to 28 days, by study, using 28-day intervals starting with the date of randomization. If a subject did not have at least 14 days of data, LOCF was used.

	Satisfying Sexual Events				Sexual Desire			
	Median change from baseline				LS Mean change from baseline*			
	Placebo	FLI 100 q h.s.	Treatment Difference	p-value**	Placebo	FLI 100 q h.s.	Treatment Difference	p-value***
Study 511.71	0.0	1.0	1.0	<0.05	6.4	8.7	2.2	>0.05
Study 511.75	0.0	1.0	1.0	<0.05	7.1	8.2	1.1	>0.05

***** Least-squared means, estimated from analysis of covariance models
****** Wilcoxon rank sum test
*** Analysis of covariance, adjusting for center and baseline
Note: Because Hochberg's method was used to adjust for multiple treatment comparisons, the p-values are reported as >0.05 or <0.05.

Source: Table D.4.1:1 and Table D.4.2:1, '*Response to Information Request (Response dated 5/14/2010, Sequence 0022)*'

Table 8. Results for the co-primary endpoints not standardized to 28 days, by study, using 28-day intervals starting with the date of randomization. All available data for a subject were used, even if there were less than 14 days of data.

	Satisfying Sexual Events				Sexual Desire			
	Median change from baseline				LS Mean change from baseline*			
	Placebo	FLI 100 q h.s.	Treatment Difference	p-value**	Placebo	FLI 100 q h.s.	Treatment Difference	p-value***
Study 511.71	0.0	1.0	1.0	<0.05	6.6	8.6	1.9	>0.05
Study 511.75	0.0	0.8	0.8	>0.05	6.9	8.5	1.6	>0.05

***** Least-squared means, estimated from analysis of covariance models
****** Wilcoxon rank sum test
*** Analysis of covariance, adjusting for center and baseline
Note: Because Hochberg's method was used to adjust for multiple treatment comparisons, the p-values are reported as >0.05 or <0.05.

Source: Table D.4.1:2 and Table D.4.2:2, '*Response to Information Request (Response dated 5/14/2010, Sequence 0022)*'

Noncompliance on eDiary Desire

Because non-compliance on the endpoint 'eDiary Desire' is possibly informative, I requested the Applicant to explore the relationship between the outcomes on 'eDiary Desire' (not reported, no desire, low desire, moderate desire, strong desire) with SSE. The analyses calculated the average SSE for each level of eDiary sexual desire by 28 day periods, starting with the date of randomization.

The results show that the average SSEs for subjects who did not complete the 'eDiary Desire' item fell between the average SSEs for subjects who reported 'no desire' and those who reported 'low desire'. This pattern was observed at Baseline and at Weeks 4, 8, 12, 16, 20 and 24; for illustration, Table 9 shows the results at Week 24.

Table 9. Average number of SSEs per day for eDiary Desire at Week 24, including subjects who did not report a level of desire

	Level of Desire at Week 24				
	Missing	No desire	Low desire	Moderate desire	Strong desire
Study 511.71	.14	.07	.16	.30	.53
Study 511.75	.14	.08	.15	.34	.51

Note: The table entries, average number of SSEs, are reported as an average for a given day. They are not normalized over 28 days.

Source: 'Response to April 29, 2010 Statistical Information Request' and 'Response to June 7 Request for Statistical Information'

An interpretation of these findings is that subjects who chose not to complete the 'eDiary Desire' item were likely to have had little or no sexual desire.

Responder analyses

An anchor-based approach was used to calculate thresholds that were used to define "response" for certain endpoints. At clinic visits, subjects completed the Patient Global Impression of Improvement (PGI-I) in order to assessed their overall improvement of their HSDD relative to the start of the study. The PGI-I is an ordinal scale ranging from 1 ("very much improved") to 7 ("very much worse"). Subjects completed the PGI-I at the end of their clinic visits, which included physical exams and the completion of other patient-reported outcome instruments.

The PGI-I is reproduced here:

Figure 13. Patient Global Index of Improvement

PGI of Improvement

1. How is your condition - meaning decreased sexual desire and feeling bothered by it - today compared to when you started study medication?

1 ☐ Very much improved

2 ☐ Much improved

3 ☐ Minimally improved

4 ☐ No change

5 ☐ Minimally worse

6 ☐ Much worse

7 ☐ Very much worse

Source: Protocol for Study 511.71

To obtain the threshold values used to define a responder, the mean changes from baseline for the endpoints shown in Table 10 were calculated for each of the seven response categories for PGI-I.

The difference between the means for 'minimally improved' and 'no change' constituted a responder. The medical division, however, had advised the Applicant that the difference between 'minimally improved' and 'much improved' should constitute a responder. Analyses comparing the proportions of responders yielded the same results, regardless of the definition used.

The results show differences between the flibanserin and placebo treatment groups for numerous endpoints, leading the Applicant to conclude that the differences for SSE and desire, as measured by FSFI-desire, were clinically meaningful; see Table 11.

However, these analyses focus only on those subjects who exceed a certain threshold. Analyses did not examine subjects whose condition worsened over time. As discussed earlier in this

section, around 30 to 40% of all subjects had decreases in SSEs and slightly more had decreases in sexual desire as assessed by the eDiary.

In addition, a different anchor-based approach to defining responders is to examine the change from baseline within one of the response categories of PGI-I – for example, the mean SSE change from baseline among subjects reporting 'minimally improved' on the PGI-I – and using this value to define a responder. A rationale for the chosen method was not provided.

Table 10. Thresholds used to define responders for selected endpoints.

Table 2.1.2.4: 1 PGI-I anchoring of select endpoints – Trials 511.70, 511.71, 511.75, 511.77 and pooled (FAS, LOCF)

	Difference between Minimally improved and No change				
Endpoint	511.70	511.71	511.75	511.77	Pooled
SSE	1.55	1.22	1.25	1.55	1.24
eDiary sexual desire score	8.15	7.80	7.91	7.25	7.87
FSFI desire items	0.83	0.83	0.74	0.88	0.78
FSDS-R Q 13	-0.47	-0.44	-0.41	-0.51	-0.42
FSDS-R total score	-4.89	-5.63	-5.07	-5.86	-5.27
FSFI total score	4.64	4.52	3.87	5.66	4.10

Source: Applicant's briefing document, dated 6/18/2010, page 35

Table 11. Results of responder analyses

	North America								
	511.71			511.75			Pooled 511.71/.75		
Endpoint	FLI 100 q.h.s. %	PBO %	Diff	FLI 100 q.h.s. %	PBO %	Diff	FLI 100 q.h.s. %	PBO %	Diff
PGI-I anchored responder endpoints									
SSE	47.6	33.0	14.7 **	44.2	34.1	10.1 *	45.7	33.6	12.0 **
eDiary sexual desire score	41.1	38.2	2.8	38.0	32.0	6.0	39.3	34.7	4.6
FSFI desire items	42.9	30.7	12.2 **	45.6	32.7	12.9 **	44.5	31.9	12.6 **
FSFI total	46.8	31.5	15.3 **	48.3	35.9	12.4 **	47.6	34.0	13.6 **
FSDS-R Q13	54.6	42.9	11.7 **	49.5	40.1	9.4 *	51.7	41.3	10.4 **
FSDS-R total	58.6	47.8	10.8 *	56.6	45.2	11.3 **	57.4	46.3	11.1 **

Source: Applicant's briefing document, dated 6/18/2010, page 74

3.2 Evaluation of Safety

Please see the medical officer's review of safety issues.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The Applicant analyzed the co-primary endpoints for numerous subgroups defined by race and ethnicity. The sample sizes were too small to allow for meaningful comparisons.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

From a statistical perspective, the evidence submitted with this application does not support the efficacy of flibanserin 100 mg q.h.s. for the treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women. I base my conclusion primarily on the finding that the two major Phase 3 studies, Study 511.71 and Study 511.75, failed to achieve statistical significance on both pre-specified co-primary endpoints. The treatment difference between flibanserin 100 mg q.h.s. and placebo was statistically significant in both studies for sexually satisfying events. However, the treatment difference for sexual desire as measured by the daily electronic diary was not statistically significant in either study.

The application indicates the eDiary measure of sexual desire was not adequately validated and suggests the lack of validation along with the frequency of noncompliance for this item may partially explain its non-significant results in the two studies. Noncompliance for the eDiary sexual desire endpoint, however, appears to provide information on a subject's level of desire and should not be treated as missing data. For example, analyses show subjects who did not respond to the eDiary desire time had a mean number of sexually satisfying sexual events (SSEs) that fell between the mean number of SSEs for subjects reporting no sexual desire and those reporting minimal sexual desire. This finding suggests that a missing value on the eDiary sexual desire item likely represents a subject whose level of desire was, at most, minimal as measured by the eDiary.

The submission proposes replacing the eDiary sexual desire item with the Female Sexual Function Index – Desire Items (FSFI-desire), which the submission claims is a better instrument for assessing sexual desire and which does not have an issue with noncompliance, because it is administered in the clinic. The Applicant's background materials for the advisory committee meeting that was held on June 18, 2010 indicate that limitations of the eDiary Desire item became evident as the clinical program progressed. Information contained in the BLA, however, is insufficient to conclude the FSFI-desire index is validated for use in a clinical trial enrolling this patient population.

If FSFI-desire is used in place of eDiary Desire as the co-primary endpoint measuring sexual desire, the results are statistically significant in both studies. From a statistical perspective, however, in addition to the lack of information supporting the validation of FSFI-desire, the use of FSFI-desire in place of the eDiary Desire endpoint is problematic, especially because FSFI-desire was not a prespecified co-primary endpoint. Although there can be situations where an endpoint other than the pre-specified endpoint could be used to assess the efficacy of a product, the substitution of a patient-reported outcome as a primary endpoint after the data are unblinded does not seem to be such a situation. In addition, the FSFI was one of a series of tests that were administered during clinic visits. The submission does not address the effects that completing these other tests may have had on a subject's responses to the FSFI and on the content validity of the FSFI.

Exploratory analyses of eDiary Desire suggest the difference at 24 weeks between flibanserin 100 mg q.h.s. and placebo depended on the baseline value of eDiary Desire. The observed treatment difference was greatest among subjects whose eDiary Desire value at baseline was relatively low. The observed treatment difference decreased with increasing values of eDiary Desire at baseline. I recommend the Applicant doing additional analyses that explore the relationship between the baseline value of eDiary Desire and the difference between treatments at 24 weeks. If these exploratory analyses suggest the possibility of a treatment effect among subjects who have no or minimal desire, the Applicant may wish to consider limiting enrollment to these subjects in future studies of flibanserin for the treatment of HSDD.

Although the protocols called for analyses of change between the 4-week baseline and the interval of Weeks 21 to 24, the analyses of the co-primary endpoints did not adhere to this plan. Instead, the eDiary data were divided into 4-week intervals that were anchored to a subject's clinic visits, and the eDiary data for the co-primary endpoints were computed as a 28-day average of all data collected between visits. As expected, clinic visits did not necessarily occur every four weeks; visits included a window of ±7 days. Moreover, the primary analyses required a minimum of 14 days of data. If less than 14 days of data were available, data were imputed by last observation carried forward (LOCF). The results of analyses that did not impute for missing data, which used all available data, were consistent with the results of the analyses that used LOCF.

Using 28-day intervals that started with the day of randomization, the Applicant repeated the analyses of the co-primary endpoints, both for the endpoints standardized to 28 days and for the actual values observed. For the SSE endpoint, the results are not as strong as the results reported for the windows anchored to clinic visits. Although the results for the data standardized to 28 days are consistent with the original analyses, the treatment difference for SSEs in Study 511.75 moves closer to statistical non-significance ($p=0.048$, adjusted for multiplicity by the Hochberg method). In the analyses of all available data, the treatment difference for SSEs is smaller and is statistically non-significant in Study 511.75.

5.2 Conclusions and Recommendations

From a statistical perspective, the evidence submitted with this application does not support the efficacy of flibanserin 100 mg q.h.s. for the treatment of hypoactive sexual desire disorder (HSDD) in premenopausal women. I base my conclusion primarily on the finding that the two major Phase 3 studies, Study 511.71 and Study 511.75, failed to achieve statistical significance on both pre-specified co-primary endpoints. The treatment difference between flibanserin 100 mg q.h.s. and placebo was statistically significant in both studies for sexually satisfying events. However, the treatment difference for sexual desire as measured by the daily electronic diary was not statistically significant in either study.

Exploratory analyses of eDiary Desire suggest the difference at 24 weeks between flibanserin 100 mg q.h.s. and placebo depended on the baseline value of eDiary Desire. The observed treatment difference was greatest among subjects whose eDiary Desire value at baseline was relatively low. The observed treatment difference decreased with increasing values of eDiary Desire at baseline. I recommend the Applicant doing additional analyses that explore the relationship between the baseline value of eDiary Desire and the difference between treatments at 24 weeks. If these exploratory analyses suggest the possibility of a treatment effect among subjects who have no or minimal desire, the Applicant may wish to consider limiting enrollment to these subjects in future studies of flibanserin for the treatment of HSDD.

In future trials of flibanserin for the treatment of HSDD, subjects who are considered non-compliant should be allowed to receive study medication and to continue in the study in order to allow analyses that more closely follow the intent-to-treat principle. Continuation in the study will minimize the amount of missing data and the resulting need for data imputation. In addition, 28-day windows should be defined based on the date of randomization and should not be tied to clinic visits.

Appendix I

The eDiary

Appendix II

Additional tables and figures

Figure 14. SSEs: Change from baseline, by study and baseline SSE grouped by quartiles.

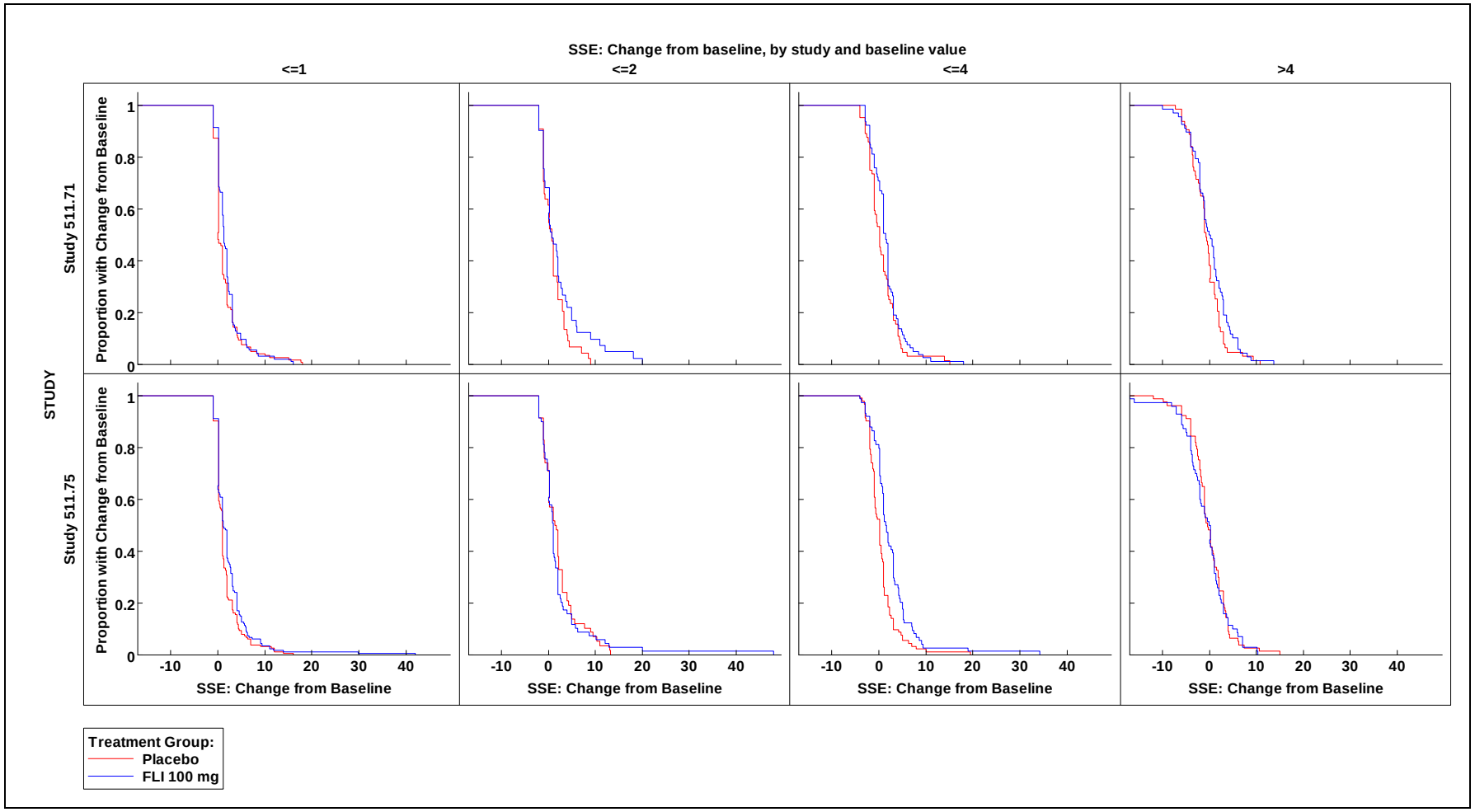
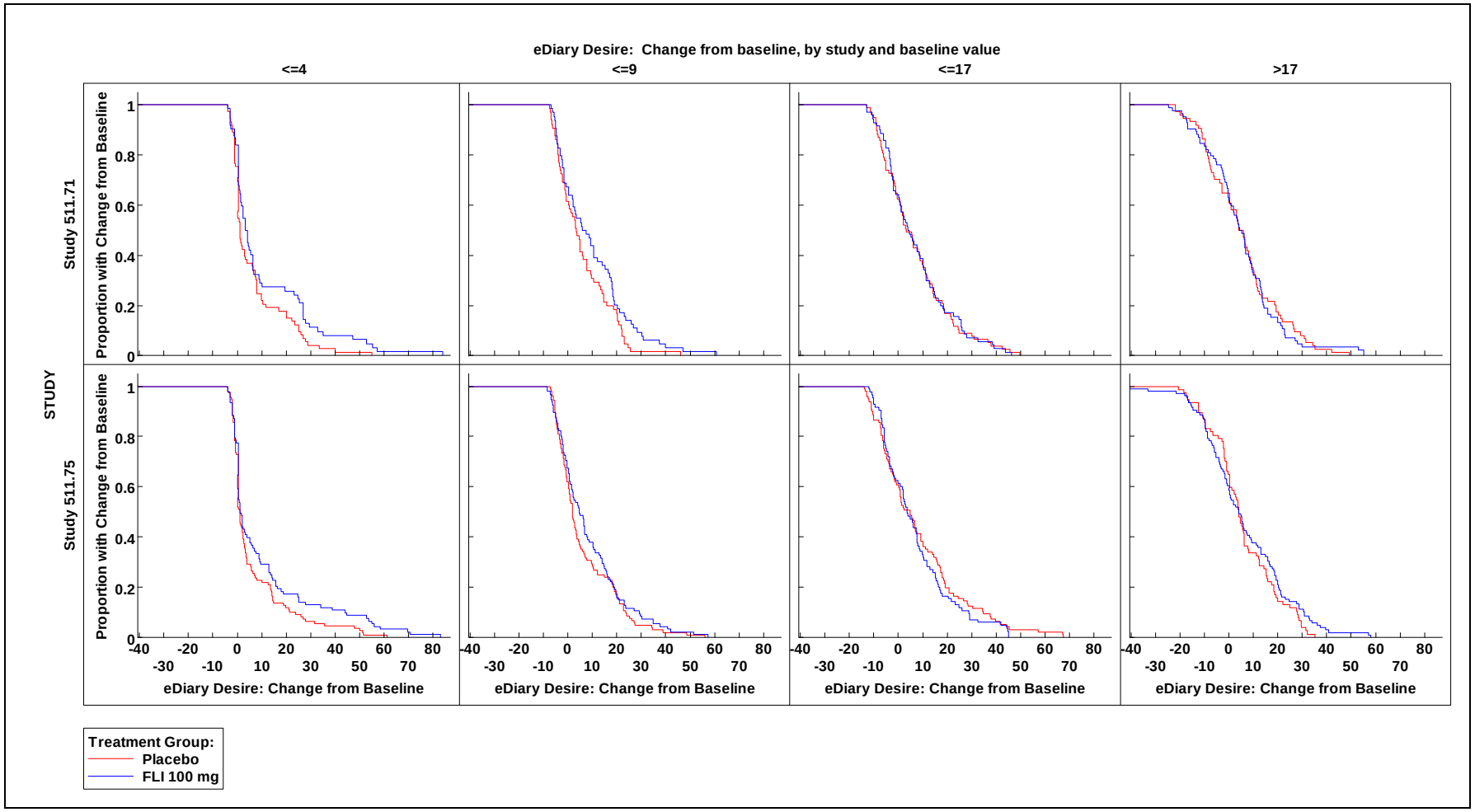


Figure 15. eDiary Desire: Change from baseline, by study and baseline eDiary Desire grouped by quartiles.



Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22526	ORIG-1	BOEHRINGER INGELHEIM PHARMACEUTICALS INC	FLIBANSERIN

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

/s/

LISA A KAMMERMAN
08/26/2010

MAHBOOB SOBHAN
08/26/2010