CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

200603

STATISTICAL REVIEW(S)

ADDENDUM

NDA/Serial Number 200603 / S 000
Drug Name: Lurasidone Hydrochloride
Indication: Schizophrenia
Applicant: Dainippon Sumitomo Pharma

This Addendum corrects a few typos appearing in my review signed off September 07, 2010. The corrections are shown in red color. These will not affect the review conclusion.

Section 3.1.2.5 Results of Efficacy Analysis

Table 25 of the original review p. 26 (below):

Table 25. Study D1050231 PANSS Total score LS Mean Change from Baseline to Week 6 (LOCF ANCOVA)

	Lurasidone 40 mg	Lurasidone 120 mg	Olanzapine 15 mg	Placebo
No patients	118	118	121	114
LS Mean change from baseline (SE)	-23.1 (1.7)	-20.0 (1.7)	-26.7 (1.7)	-15.2 (1.7)
Placebo-adjusted Difference	-7.9 (2.4)	-4.8 (2.4)	-11.4 (2.4)	NA
p-value	0.001	0.049	<0.001	NA
Adjusted p-value	0.001 (Hommel)	0.098 (Hommel)		

Source: Table 19 Clinical Study Report D1050231 (p. 86)

should be replaced by:

Table 25. Study D1050231 PANSS Total score LS Mean Change from Baseline to Week 6 (LOCF ANCOVA)

	Lurasidone 40 mg	Lurasidone 120 mg	Olanzapine 15 mg	Placebo
No patients	118	118	121	114
LS Mean change from baseline (SE)	-23.1 (1.7)	-20.0 (1.7)	-26.7 (1.7)	-15.2 (1.7)
Placebo-adjusted Difference	-7.9 (2.4)	-4.8 (2.4)	-11.4 (2.4)	NA
p-value	0.001	0.049	<0.001	NA
Adjusted p-value	**0.002** (Hommel)	0.098 (Hommel)		

Source: Table 19 Clinical Study Report D1050231 (p. 86)

Section 5.1 Statistical Issues and Collective Evidence

Table 33 of the original review p. 31 (below):

Table 33. Summary of Efficacy Results for The Primary Endpoints

		LS Mean Treatment Difference from Placebo			
Study	**Primary Endpoint**	**Lurasidone 40 mg**	**Lurasidone 80 mg**	**Lurasidone 120 mg**	**Olanzapine 15 mg**
D1050006	BPRS (ANCOVA)	-5.6 p=0.018 (Dunnett)	NA	-6.7 p=0.004 (Dunnett)	NA
D1050196	BPRS (ANCOVA)	NA	-4.7 p=0.0118	NA	NA
D1050229	PANSS (MMRM)	-2.1 p=0.591 (Hommel-based)	-6.4 p=0.034 (Hommel-based)	-3.5 p=0.391 (Hommel-based)	NA
D1050231	PANSS (MMRM)	-9.7 p=0.02 (Hommel-based)	NA	-7.5 p=0.022 (Hommel-based)	-12.6 P<0.001

Source: Clinical study reports D1050006, D1050196, D1050229, D1050231; p-values were adjusted using pre-specified multiple testing procedures: Dunnett procedure for Study D1050006; the Hommel-based gatekeeping procedure for studies D1050229, D1050231

should be replaced by:

Table 33. Summary of Efficacy Results for The Primary Endpoints

		LS Mean Treatment Difference from Placebo			
Study	**Primary Endpoint**	**Lurasidone 40 mg**	**Lurasidone 80 mg**	**Lurasidone 120 mg**	**Olanzapine 15 mg**
D1050006	BPRS (ANCOVA)	-5.6 (2.13) p=0.018 (Dunnett)	NA	-6.7 (2.16) p=0.004 (Dunnett)	NA
D1050196	BPRS (ANCOVA)	NA	-4.7 (1.84) p=0.0118	NA	NA
D1050229	PANSS (MMRM)	-2.1 (2.5) p=0.591 (Hommel-based)	-6.4 (2.5) p=0.034 (Hommel-based)	-3.5 (2.5) p=0.391 (Hommel-based)	NA
D1050231	PANSS (MMRM)	-9.7 (2.9) **p=0.002** (Hommel-based)	NA	-7.5 (3.0) p=0.022 (Hommel-based)	-12.6 (2.8) P<0.001

Source: Clinical study reports D1050006, D1050196, D1050229, D1050231; p-values were adjusted using pre-specified multiple testing procedures: Dunnett procedure for Study D1050006; the Hommel-based gatekeeping procedure for studies D1050229, D1050231

Table 35 of the original review p. 32 (below):

Table 35. Summary of Efficacy Results for Change from Baseline in PANSS Total Score Based on MMRM Analysis.

Study	Primary Endpoint	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Olanzapine 15 mg
D1050006	PANSS (MMRM)	-14.1 p=0.009 (unadjusted)	NA	-16.2 p=0.0027 (unadjusted)	NA
D1050196	PANSS (MMRM)	NA	-10.2 p=0.0064	NA	NA
D1050229	PANSS (MMRM)	-2.1 p=0.591 (Hommel-based)	-6.4 p=0.034 (Hommel-based)	-3.5 p=0.391 (Hommel-based)	NA
D1050231	PANSS (MMRM)	-9.7 p=0.02 (Hommel-based)	NA	-7.5 p=0.022 (Hommel-based)	-12.6 p<0.001

Source: Clinical study reports D1050229 and D1050231, and Reviewer's results

should be replaced by:

Table 35. Summary of Efficacy Results for Change from Baseline in PANSS Total Score Based on MMRM Analysis.

Study	Primary Endpoint	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Olanzapine 15 mg
D1050006	PANSS (MMRM)	-14.1 p=0.009 (unadjusted)	NA	-16.2 p=0.0027 (unadjusted)	NA
D1050196	PANSS (MMRM)	NA	-10.2 p=0.0064	NA	NA
D1050229	PANSS (MMRM)	-2.1 p=0.591 (Hommel-based)	-6.4 p=0.034 (Hommel-based)	-3.5 p=0.391 (Hommel-based)	NA
D1050231	PANSS (MMRM)	-9.7 **p=0.002** (Hommel-based)	NA	-7.5 p=0.022 (Hommel-based)	-12.6 p<0.001

Source: Clinical study reports D1050229 and D1050231, and Reviewer's results

---

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

/s/

GEORGE KORDZAKHIA
09/24/2010

PEILING YANG
09/24/2010
I agree that these typos have no impact on the review conclusions.

KOOROS MAHJOOB
09/24/2010
I concure.

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/Serial Number:	200603/ S 000
Drug Name:	Lurasidone Hydrochloride
Indication(s):	Schizophrenia
Applicant:	Dainippon Sumitomo Pharma
Date(s):	Initial submission date: December 30, 2009
Review Priority:	Standard
Biometrics Division:	Division of Biometrics I
Statistical Reviewer:	George Kordzakhia, Ph.D.
Concurring Reviewers:	Peiling Yang, Ph.D; Kooros Mahjoob, Ph.D.
Medical Division:	Division of Psychiatry Products
Clinical Team:	Cara Alfaro, Pharm. D., Reviewer Ni Khin , M.D., Team Leader
Project Manager:	Ann Sohn, Pharm. D.

Keywords: clinical studies, NDA review

TABLE of CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

1.1 CONCLUSIONS AND RECOMMENDATIONS

The sponsor submitted results of four studies (D1050006, D1050196, D1050229, D1050231) in support of efficacy of lurasidone (fixed doses of 40 mg/day, 80 mg/day, and 120 mg/day) versus placebo for the treatment of schizophrenia (b) (4).

In the primary analysis of BPRS Total score (studies D1050006, D1050196) and PANSS Total score (studies D1050229, D1050231), adult patients with a primary diagnosis of schizophrenia on lurasidone (fixed doses of 40 mg, 80 mg, and 120 mg) were observed to show statistically significant improvement over patients in the placebo treatment group. However, the 120 mg dose did not seem to add additional benefit over the other two doses. The results from study D1050006 with 70% dropout rate and the results from study D1050196 with 50% dropout rate should be interpreted with extra caution.

CGI-S was the pre-specified key secondary endpoint in studies D1050229 and D1050231. Study D1050229 demonstrated a statistical significance with respect to this endpoint for dose 80 mg, and study D1050231 for both doses 40 mg and 120 mg. CGI-S was declared as a secondary endpoint, but not pre-specified as a key secondary endpoint, in studies D1050006 and D1050196. Nevertheless, in study D1050006 the p-values from both doses (40 mg and 120 mg) were very small so that any multiple testing procedure would lead to a statistical significance for both doses. Study D1050196 investigated 80 mg only, and the p-value from testing this endpoint was also very small. Overall, lurasidone in doses 40 mg, 80 mg, and 120 mg was statistically superior to placebo in change from baseline in CGI-S score at week 6. However, whether the magnitude of improvement was of clinical relevancy is deferred to the clinical review team.

1.2 BRIEF OVERVIEW OF CLINICAL STUDIES

The effectiveness of lurasidone in the treatment of schizophrenia was investigated in five, randomized, double-blind, placebo-controlled, 6-week, multicenter studies: Studies D1050006, D1050049, D1050196, D1050229, and D1050231. These studies evaluated subjects with a primary diagnosis of schizophrenia who had an acute exacerbation of psychotic symptoms.

Study D1050049 failed to show efficacy versus placebo for any lurasidone dose group or the active comparator, haloperidol. Thus, only studies D1050006, D1050196, D1050229, and D1050231 were considered for efficacy evaluation. All four studies had fixed-dose administration of lurasidone at the target therapeutic doses (40 mg, 80 mg, and 120 mg) over a period of 6 weeks. Study D1050231 also included an active comparator (olanzapine) in order to assess study assay sensitivity. These studies include data from 1,307 subjects who received study medication (800 on lurasidone, 384 on placebo, and 123 on olanzapine). Lurasidone was assessed at once daily doses of 40 mg and 120 mg in Studies D1050006 and D1050231, 80 mg in Study D1050196, and 40 mg, 80 mg and 120 mg in Study D1050229.

1.3 STATISTICAL ISSUES AND FINDINGS

Studies D1050006 and D1050196

In Study D1050006, the sponsor investigated efficacy of SM-13496 40 mg and SM-13496 120 mg. Study D1050006 had approximately 70% dropout rate. Based on LOCF ANCOVA analysis, SM-13496 (Lurasidone) treatment groups were statistically superior to placebo in mean change from baseline to Week 6 in BPRS Total score (primary endpoint). The Dunnett's adjusted p-values of pairwise comparisons with placebo were 0.018 (SM-13496 40 mg versus Placebo) and 0.004 (SM-13496 120 mg versus Placebo). However, the 120 mg dose did not seem to add additional benefit over the 40 mg dose. In addition, the strength of evidence may be weakened by the considerably high dropout rate along with the relatively small sample size in each group. The efficacy results in Study D1050006 should be interpreted with extra caution.

Study D1050196 had two treatment arms: SM-13496 80 mg and placebo. The study had 50% dropout rate. Based on LOCF ANCOVA analysis, SM-13496 80 mg was statistically significantly better than placebo in mean change from baseline to Week 6 in BPRS total score (p-value 0.0118). The strength of evidence may be weakened by the considerably high dropout rate. Extra caution should be made for any interpretation of the results.

In both studies, change from baseline in CGI-S score was a secondary endpoint, but not a pre-specified key secondary endpoint. No multiplicity adjustment for null hypotheses associated with CGI-S was prespecified in study D1050006. Nevertheless, in study D1050006 the LOCF ANCOVA p-values from both doses (40 mg and 120 mg) were very small so that any multiple testing procedure would lead to a statistical significance for both doses. Study D1050196 investigated 80 mg only, and the p-value from testing this endpoint was also very small. Whether the magnitude of improvement in CGI-S was of clinical relevancy is deferred to the clinical review team.

Studies D1050229 and D1050231

In Study D1050229 three doses of lurasidone (40 mg, 80 mg, and 120 mg) were compared with placebo with respect to the primary endpoint (change from baseline in PANSS) and the key secondary endpoint (change from baseline in CGI-S). After multiplicity adjustment using the Hommel-based gatekeeping procedure with truncation parameter gamma=0.5, lurasidone 80 mg was statistically significantly better than placebo in both, the primary and the key secondary endpoints. Doses of 40 mg and 120 mg failed to demonstrate efficacy for either of the two endpoints.

Study D1050231 included four treatment arms: two lurasidone treatment arms (40 mg and 120 mg), placebo, and active comparator (Olanzapine 15 mg). Numerically, Olanzapine 15 mg showed better treatment effect compared to lurasidone and placebo arms. After multiplicity adjustment, using the Hommel-based gatekeeping procedure with truncation parameter zero, both lurasidone treatment arms (40 mg and 120 mg) were statistically significantly superior to placebo in both, the primary (change from baseline in PANSS) and the key secondary (change from baseline in CGI-S) endpoints.

Whether the magnitudes of improvement in CGI-S in studies D1050229 and D1050231 were of clinical relevancy is deferred to the clinical review team.

2 INTRODUCTION

2.1 OVERVIEW

The effectiveness of lurasidone in the treatment of schizophrenia was investigated in five, randomized, double-blind, placebo-controlled, 6-week, multicenter studies: Studies D1050006, D1050049, D1050196, D1050229, and D1050231. These studies evaluated subjects with a primary diagnosis of schizophrenia who had an acute exacerbation of psychotic symptoms.

Study D1050049 failed to show efficacy over placebo for any lurasidone dose group (20 mg, 40 mg, or 80 mg) or the active comparator, haloperidol 10 mg. Thus, only studies D1050006, D1050196, D1050229, and D1050231 were considered in this statistical review. All four studies had fixed-dose administration of lurasidone at the target therapeutic doses (40 mg, 80 mg, and 120 mg) over a period of 6 weeks. Study D1050231 also included an active comparator arm (olanzapine) in order to assess study assay sensitivity.

2.2 DATA SOURCES

Data used for review are from the electronic submissions received on January 4, 2010. The network path is \\Cdsesub1\evsprod\NDA200603\0000.

3 STATISTICAL EVALUATION

3.1 EVALUATION OF EFFICACY

3.1.1 STUDY D1050006 AND STUDY D1050196

3.1.1.1 Objective

The primary objective of studies D1050006 and D1050196 was to evaluate the efficacy of SM-13496 versus placebo in the treatment of subjects with acute exacerbation of schizophrenia, as measured by reductions from baseline in total score of the Brief Psychiatric Rating Scale (BPRS), as extracted from the Positive and Negative Symptom Scale (PANSS). Study D1050006 evaluated the efficacy of SM-13496 at 40 mg/day and 120 mg/day, and study D1050196 evaluated the efficacy of SM-13496 at 80 mg/day.

3.1.1.2 Study Design

Both studies were 6-week, multicenter, randomized, fixed-dose, double-blind, parallel-group, placebo-controlled clinical trials. The studies included 3 periods: a screening period (up to 14 days), a single-blind placebo washout period (3-7 days), and a double-blind treatment period (6 weeks). Following the washout period (inpatient), subjects meeting entry criteria were randomly assigned to SM-13496 40 mg or 120 mg, or placebo (Study D1050006), and SM-13496 80 mg, or placebo (Study D1050196), given in a once-daily regimen for 6 weeks.

Eleven visits were scheduled: 1 to screen candidate subjects (Visit 1, Screening), 1 to dispense single-blind placebo medication (Visit 2, Washout), 1 to provide baseline assessments (Visit 3, Baseline), 1 to give the first double-blind treatment (Visit 4, Day 1), and 7 to assess double-blind treatment (Visits 5, 6, 7, 8, 9, 10, and 11).

Table 1. Study Flow Chart (Studies D1050006 and D1050196)

Period	**Screening**	**Washout**	**Baseline**	**Double-Blind Treatment**
Day	-20 to -7	-6 to -1	0	1, 3, 7, 14, 21, 28, 35, 42
Visit	Visit1	Visit 2	Visit 3	Visit 4, 5, 6, 7, 8, 9, 10, 11

3.1.1.3 Patient Disposition, Demographic and Baseline Characteristics

Study D1050006

This was a multicenter study at 15 sites in the U.S. The study was conducted over a period of 10 months. The first subject was enrolled on February 6, 2001, and the last subject completed the study on December 18, 2001.

Overall, 149 subjects were randomized to study medication (50 to SM-13496 40 mg, 49 to SM-13496 120 mg, and 50 to placebo), of which 98 discontinued and 51 completed. Subject evaluation groups (Safety and ITT populations) and reasons for discontinuation are summarized in Table 2. Overall, the most common reason for discontinuation was withdrawal of consent (24.8% of subjects). Lack of efficacy was the most frequent reason for discontinuation in the placebo group (32.0%). The 120 mg group had the lowest overall discontinuation rate (59.2%). The placebo group had the lowest discontinuation rate for AEs, while there was no difference between the two doses of the investigational drug.

Table 2. Study D1050006 Patient Population and Disposition

Patients	**Placebo**	**SM-13496 40 mg**	**SM-13496 120 mg**
Randomized	50 (100%)	50 (100%)	49 (100%)
Safety	50 (100%)	50 (100%)	49 (100%)
ITT Population	49 (98.0%)	49 (98%)	47 (95.9%)
Discontinued Study	35 (70.0%)	34 (68.0%)	29 (59.2%)
Luck of Efficacy	16 (32.0%)	11 (22.0%)	6 (12.2%)
Withdrawal of Consent	11 (22.0%)	13 (26.0%)	13 (26.5%)
Adverse Event	2 (4.0%)	6 (12.0%)	6 (12.2%)
Protocol Violation	2 (4.0%)	1 (2.0%)	0 (0.0%)
Lost to Follow-up	2 (4.0%)	2 (4.0%)	2 (4.1%)
Other	2 (4.0%)	1 (2.0%)	2 (4.1%)
Completed study	15 (30%)	16 (32%)	20 (40.8%)

Source: Table 7.1.1.(pg. 44) Clinical Study Report D1050006

Table 3 summarizes demographic and baseline characteristics of the safety population. As shown in Table 3, the majority of subjects were male (84.0%, 72.0%, and 73.5% in the placebo, SM-13496 40 mg, and SM-13496 120 mg groups, respectively). Subjects in the three groups were comparable in age (mean age of 38.1, 39.8, and 41.0 years for the placebo, SM-13496 40 mg, and SM-13496 120 mg groups, respectively. Seventy-four (49.7%) subjects were Black, 62 (41.6%) were White, 7 (4.7%) were Hispanic, 2 (1.3%) were Asian, and 4 (2.7%) were “Other”. The three treatment groups were comparable in almost all of the above demographic and baseline characteristics (BPRS Total score at baseline).

Table 3. D1050006 Demographic and Baseline Characteristics (Safety Population)

Variable	Placebo N=50	SM-13496 40 mg N=50	SM-13496 120 mg N=49
Male	42 (84%)	36 (72%)	36 (73.5%)
Female	8 (16%)	14 (28%)	13 (26.5%)
Age (years)			
Mean (min, max)	38.1 (18, 56)	39.8 (21, 61)	41.0 (24, 59)
Race			
Caucasian	20 (40%)	20 (40%)	22 (44.9%)
Black	25 (50%)	25 (50%)	24 (49.0%)
Asian	1 (2.0%)	1 (2.0%)	0 (0.0%)
Hispanic	1 (2.0%)	3 (6.0%)	3 (6.1%)
Other	3 (6.0%)	1 (2.0%)	0 (0.0%)
BPRS Total Score			
Mean (SD)	54.4 (8.3)	54.6 (9.1)	52.5 (7.6)

Source: Table 7.5.1.(pg. 46), Table 7.6.1. (pg. 47) Clinical Study Report D1050006

Study D10500196

The clinical study was conducted at 22 sites in the United States. The first subject was randomized on May 28, 2004, and the last subject completed the study on December 6, 2004.

Overall, 180 subjects were randomized and 180 subjects received study medication (90 received placebo, 90 received SM-13496 80 mg), of whom 81 discontinued and 99 completed. Subject populations (Safety, ITT) and reasons for discontinuation are summarized in Table 4. Overall, the most common reasons for discontinuation were lack of efficacy (21.1%) and withdrawal of consent (15.0%). The overall discontinuation rate was slightly higher in the placebo group (47.8%) compared with the SM-13496 group (42.2%).

Table 4. Study D10500196 Patient Population and Disposition

Patients	Placebo	SM-13496 80 mg
Randomized	90 (100%)	90 (100%)
Received Study Drug	90 (100%)	90 (100%)
ITT Population	90 (100%)	90 (100%)
Discontinued Study	43 (47.8%)	38 (42.2%)
Luck of Efficacy	29 (32.2%)	9 (10.0%)
Withdrawal of Consent	9 (10.0%)	18 (20.0%)
Lost to Follow-up	2 (2.2%)	1 (1.1%)
Other	2 (2.2%)	4 (4.4%)
Adverse Event	1 (1.1%)	6 (6.7%)
Protocol Violation	0	0
Death	0	0
Completed study	47 (52.2%)	52 (57.8%)

Source: Table 10.1.1. (p.50) Clinical Study Report D1050196

The two treatment groups were comparable in almost all of the demographic and baseline characteristics shown in Table 5. Of the 180 subjects, 138 (76.7%) were men. Black subjects were the largest racial group, comprising 57.2% of the population (103 of 180 subjects; 62% in the placebo group and 52% in the SM-13496 group). The predominant schizophrenia subtype was paranoid (80.6%; 145 of 180 subjects).

Table 5. Study D1050196 Demographic and Baseline Characteristics (Safety Population)

Variable	Placebo N=90	SM-13496 80 mg N=90
Gender, n (%)		
Male	70 (77.8%)	68 (75.6%)
Female	20 (22.2%)	22 (24.4%)
Race		
Caucasian	26 (28.9%)	35 (38.9%)
Black	56 (62.2%)	47 (52.2%)
Asian	1 (1.1%)	2 (2.2%)
Hispanic	7 (7.8%)	5 (5.6%)
Other	0 (0.0%)	1 (1.1%)
Age (years)		
Mean (SD)	41.9 (9.78)	39.7 (9.91)
Weight (kg)		
Mean (SD)	93.3 (22.35)	91.8 (25.61)
Height (cm)		
Mean (SD)	173.2 (9.23)	173.2 (9.11)
BPRS Total Score		
Mean (SD)	56.1 (6.84)	55.1 (5.95)
PANSS Total Score		
Mean (SD)	96.0 (11.59)	94.4 (10.90)

Source: Table 10.5.1. (p.53), Table 10.6.1. (p.54) Clinical Study Report D1050196

3.1.1.4 Statistical Methodologies

For both studies, the primary efficacy variable was the change from baseline to Day 42 in the BPRS total score (as extracted from the PANSS by adding scores on items P2-P7, N1-N2, and G1-G10, on a 1- to 7-point scale per item). The primary analysis was performed using the LOCF approach in the ITT population. The ITT population included all randomized subjects who took at least 1 dose of study medication and had at least 1 efficacy evaluation on treatment.

Primary Analysis for the Primary Endpoint

The primary analysis for testing the change from baseline to Day 42 in BPRS score was the 2-way analysis of covariance (ANCOVA) model with treatment group and study center (pooled) as main effects, and baseline BPRS score as a covariate. For study D1050006 pairwise comparisons (lurasidone 40 mg vs. placebo, and lurasidone 120 mg vs. placebo) were performed using a Dunnett's test at a 2-sided significance level of 0.05 to adjust for multiplicity.

As an exploratory analysis by the sponsor, the change from baseline to Day 42 in BPRS score was analyzed by the same ANCOVA model for the OC (Observed Cases) population set.

Pooling of Study Centers

Study D1050006: Study centers with <=5 subjects were pooled for primary efficacy analysis. Centers were combined until there were at least 2 subjects per treatment per center.

Study D1050196: Study centers not having BPRS data (baseline, and at least one post-baseline score) for at least 2 subjects per treatment group were pooled. The pooling of centers was done only for the efficacy data.

Handling of Missing Data

Study D1050006: For individual scale items of the PANSS, the last available score was carried forward. If more than 30% of individual items were missing at a particular visit, then the entire scale for that visit was designated missing. Baseline and screening items were not carried forward into the double-blind treatment phase for missing items. When baseline scores were missing for the following measures, screening scores were substituted: BPRS, total PANSS, CGI-S, AIMS, BAS, SAS.

Study D1050196: For missing individual items (questions or statements) of the BPRS and PANSS, the last available individual item that was not a baseline value was carried forward provided that less than 30% of the individual items were not missing. For all other scales (CGI-S, MADRS, AIMS, BAS, and SAS), if an individual item was missing at a given visit, the value of the missing item was carried forward from the last non-baseline visit at which the item was present. For all variables with missing baseline scores, the screening score, if available, was substituted.

Secondary Endpoint

The change from baseline in CGI-S score at Week 6 was specified as a secondary efficacy variable, but not as a key-secondary variable. The between–treatment group analysis was performed using an ANCOVA model with treatment group and study center as factors and the baseline CGI-S score as a covariate.

3.1.1.5 Results of Efficacy Analysis

Study D1050006

Primary Efficacy Analysis for The Primary Endpoint

This reviewer confirmed sponsor's primary efficacy analysis result. The primary efficacy variable was the change from baseline to Week 6 in BPRS total score. Table 6 presents the results of the ANCOVA analysis for this primary endpoint, using the LOCF approach. The changes from baseline to Week 6 in the lurasidone 40 mg and 120 mg groups were statistically significantly greater than that in the placebo group with respective Dunnett adjusted p-values 0.018 and 0.004.

Table 6. Study D1050006 BPRS Total Score LS Mean Change from Baseline to Week 6 (LOCF ANCOVA)

		Placebo	**SM-13496 40 mg**	**SM-13496 120 mg**
No patients	N=145	49	49	47
Baseline	Mean (SD)	54.7 (8.13)	54.2 (8.93)	52.7 (7.61)
Change from Baseline	LS Mean (SE)	-3.8 (1.57)	-9.4 (1.58)	-11 (1.58)
Placebo-adjusted difference	LS Mean Difference	NA	-5.6 (2.13)	-6.7 (2.16)
	95% CI	NA	(-9.8, -1.4)	(-11, -2.5)
	p-value (Dunnett)	NA	0.018	0.004

Source: Table 8.1.1. (pg. 52) Clinical Study Report D1050006

As seen from Table 7 the observed treatment difference was numerically in favor of SM-13496 at every visit during the double-blind phase.

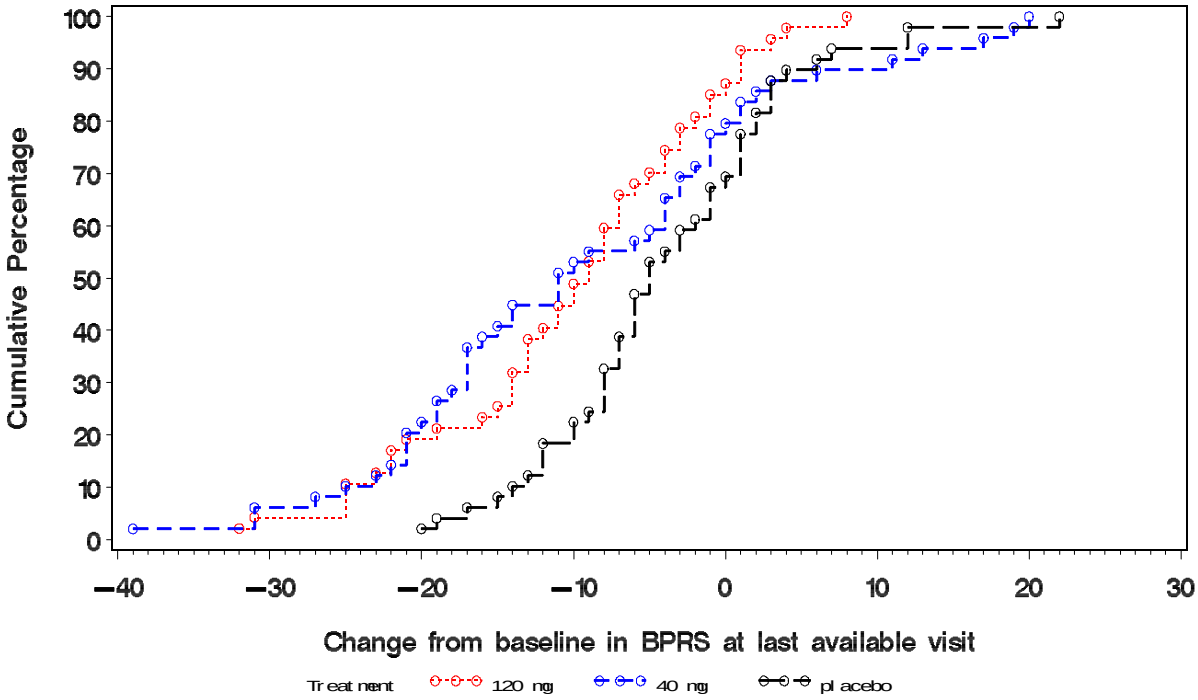
Table 7. Study D1050006 BPRS Total Score Mean Change from Baseline by Visit (LOCF ANCOVA).

	Placebo	SM-13496 40 mg	SM-13496 120 mg
Day	**LS Mean (SE)**	**LS Mean (SE)**	**LS Mean (SE)**
3	-3.0 (1.04)	-4.6 (1.01)	-5.0 (1.02)
7	-4.3 (1.33)	-5.6 (1.33)	-6.4 (1.32)
14	-4.7 (1.49)	-8.4 (1.50)	-10.4 (1.50)
21	-4.7 (1.59)	-10.1 (1.60)	-9.7 (1.60)
28	-4.9 (1.56)	-9.4 (1.57)	-10.6 (1.57)
35	-4.6 (1.55)	-9.4 (1.56)	-10.3 (1.56)

Source: Reviewers Results

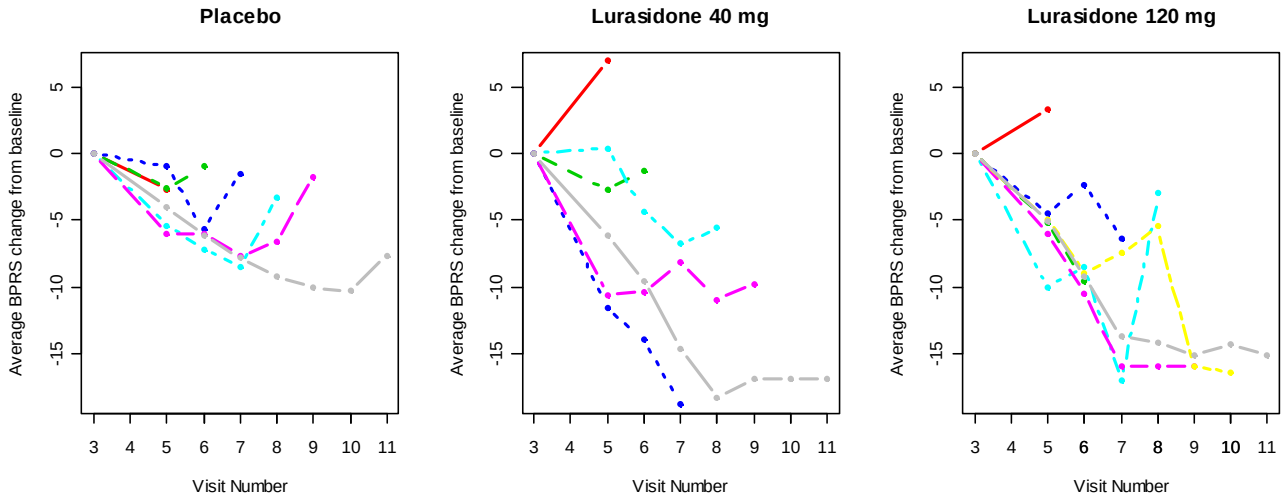
Figure 1 displays empirical cumulative distribution functions (CDF) of the primary endpoint, change from baseline in BPRS at week 6 (LOCF), for the three treatment arms in Study D1050006. Negative values of the primary endpoint represent improvement. The cumulative distribution functions describe the percentage of patients (vertical axis) in each treatment arm with primary endpoint values (horizontal axis) equal to or less than a given number x where x varies from -40 to 30. For negative values of x, the CDFs for SM-13496 treatment arms separate from the CDF for placebo arm. Numerically, in both SM-13496 arms larger proportions of patients had negative value of the primary endpoint compared with placebo arm. Because of the very high dropout rate, the graph should be interpreted with extra caution.

Figure 1. Study D1050006 Empirical cumulative distribution functions



Source: Reviewer's results

Figure 2. BPRS total score response profiles by treatment group



Source: Reviewer's results

Each curve on the plots of the Figure 2 shows the change of the BPRS total score by visit averaged over the patients, grouped by their last visit (5, 6, 7, 8, 9, 10, and 11). The plots illustrate the tendency of the patients to drop out from the study as their BPRS total score increases (getting worse). For the SM-13496 (lurasidone) treatment arms, those patients who stayed in the double-blind phase longer tend to have larger improvement from the beginning to the last visit than those who dropped out earlier. The plots provide no evidence that the missing data mechanism is MCAR (missing completely at random), which is required for the LOCF imputation approach. Although the LOCF ANCOVA is the prespecified primary analysis, the strength of evidence is weakened because of the MCAR assumption, so results have to be interpreted with caution.

Sensitivity Analysis for the Primary Endpoint

This reviewer conducted sensitivity analysis on the primary endpoint. Change from baseline in BPRS total score was analyzed by mixed model with repeated measures (MMRM). The model included study center, treatment group, visit, and treatment group-by-visit interaction as factors and baseline BPRS total score as a covariate. The findings supported the primary analysis results. Both SM-13496 treatment arms were statistically significantly better than placebo (see Table 8).

Table 8. Study D1050006 BPRS Total Score LS Mean Change from Baseline (MMRM analysis)

	Placebo		**SM-13496 40 mg**		**SM-13496 120 mg**	
Day	**No**	**LS Mean (SE)**	**No**	**LS Mean (SE)**	**No**	**LS Mean (SE)**
3	44	-3.2 (1.03)	47	-4.6 (1.01)	44	-5.0 (1.01)
7	44	-4.7 (1.37)	47	-6.2 (1.34)	44	-6.9 (1.37)
14	34	-5.5 (1.50)	32	-11.0 (1.53)	37	-11.5 (1.44)
21	26	-5.2 (2.00)	26	-13.2 (2.00)	24	-10.0 (2.01)
28	18	-6.5 (1.93)	22	-12.5 (1.87)	23	-12.8 (1.83)
35	15	-5.9 (1.92)	17	-13.2 (1.88)	21	-12.7 (1.78)
42	17	-4.1 (2.11)	17	-13.4 (2.10)	19	-13.4 (2.00)
Day 42	LS Mean Difference (SE)		-9.3 (2.95)		-9.2 (2.89)	
	Unadjusted p-value		0.0025		0.0022	

Source: Reviewer's Results

This reviewer also confirmed sponsor's OC analysis of the primary endpoint which was based on ANCOVA model with factors treatment and center, and BPRS baseline score as a covariate. After Dunnett's adjustment, neither of the two treatment placebo comparisons was significant at the 0.05 significance level (see Table 9).

Table 9. Study D1050006 BPRS Total Score Mean Change from Baseline by Visit (OC ANCOVA).

	Placebo	**SM-13496 40 mg**	**SM-13496 120 mg**
Day	**LS Mean (SE)**	**LS Mean (SE)**	**LS Mean (SE)**
3	-3.0 (1.04)	-4.6 (1.01)	-5.0 (1.02)
7	-4.6 (1.39)	-6.3 (1.34)	-7.0 (1.36)
14	-8.2 (1.35)	-14.6 (1.37)	-12.3 (1.24)
21	-10.4 (1.77)	-15.5 (1.73)	-12.70 (1.70)
28	-12.2 (1.81)	-17.5 (1.57)	-16.7 (1.50)
35	-13.2 (2.17)	-17.4 (1.88)	-15.7 (1.81)
42	-9.9 (2.4)	-16.6 (2.15)	-14.9 (2.15)
Day 42	LS Mean Difference	-6.7 (3.06)	-4.9 (2.89)
	Unadjusted p-value	0.0345	0.0950
	p-value (Dunnett)	0.062	0.164

Source: End-of-text Table 7.1 Clinical Study Report D1050006

Supportive Secondary Endpoint (PANSS)

As an exploratory analysis, change from baseline in PANSS Total score was analyzed by ANCOVA model with terms for treatment, and pooled center, and PANSS baseline score as a covariate. After Dunnett's adjustment, only SM-13496 120 mg treatment arm was statistically significantly better than placebo (see Table 10). This reviewer also considered MMRM analysis for the PANSS endpoint. Based on the MMRM analysis, both doses were statistically significantly better than placebo (see Table 10).

Table 10. Study D1050006 PANSS Total Score LS Mean Change from Baseline to Week 6

		Placebo	**SM-13496 40 mg**	**SM-13496 120 mg**
ANCOVA LOCF				
Change from Baseline	LS Mean (SE)	-6.2 (2.74)	-13.8 (2.74)	-17.0 (2.73)
Placebo-adjusted difference	LS Mean Difference	NA	-7.6 (3.67)	-10.7 (3.74)
	Unadjusted p-value	NA	0.0414	0.0047
	p-value (Dunnett)	NA	0.076	0.009
MMRM				
Change from Baseline	LS Mean (SE)	-4.6 (3.77)	-18.7 (3.75)	-20.8 (3.59)
Placebo-adjusted difference	LS Mean Difference	NA	-14.1 (5.25)	-16.2 (5.16)
	Unadjusted p-value	NA	0.0090	0.0027

Source: Table 8.2.1 (p. 52) Clinical Study Report D1050106 and Reviewer's Results

Secondary Endpoint

Table 11 presents efficacy results for the CGI-S score at Week 6, the secondary efficacy endpoint. At Week 6, statistically significant improvement was seen in the SM-13496 groups relative to the placebo group in the LOCF ANCOVA with treatment group and study centers as factors and baseline CGI-S score as a covariate (p=0.0009 [CI: (-1.1, -0.3)] and p=0.0006

[CI: (-1.1, -0.3)] for the SM-I3496 40 mg vs placebo and SM-I3496 120 vs placebo comparisons respectively). Since change from baseline in CGI-S was not prespecified as a key secondary endpoint, no multiplicity adjustment for null hypotheses associated with CGI-S was prespecified in study D1050006. Nevertheless, the LOCF ANCOVA p-values from both doses (40 mg and 120 mg) were very small so that any multiple testing procedure would lead to a statistical significance for both doses. Whether the magnitude of improvement was of clinical relevancy is deferred to the clinical review team. This reviewer verified the sponsor's results and conducted exploratory analysis using MMRM. The individual p-values for treatment-placebo comparisons were 0.038 (40 mg vs placebo) and 0.03 (120 mg vs placebo), however the p-values were not adjusted for multiplicity. It appears that CGI-S was assessed only at the last post-baseline visit for each patient (each patient had only one post-baseline CGI-S score), so MMRM results cannot be used to support efficacy.

Table 11. Study D1050006 CGI-S Change from Baseline to Week 6 (LOCF ANCOVA)

		Placebo	**SM-13496 40 mg**	**SM-13496 120 mg**
Baseline	Mean (SD)	4.7 (0.66)	4.8 (0.72)	4.7 (0.62)
ANCOVA				
Change from Baseline	LS Mean (SEM)	-0.1 (0.14)	-0.8 (0.15)	-0.8 (0.14)
Placebo-adjusted difference	LS Mean Difference (SEM)	NA	-0.7 (0.20)	-0.7 (0.20)
	95% CI	NA	(-1.1, -0.3)	(-1.1, -0.3)
	p-value	NA	0.0009	0.0006
MMRM				
Change from Baseline	LS Mean (SEM)	-0.4 (0.25)	-1.1 (0.24)	-1.1 (0.22)
Placebo-adjusted difference	LS Mean Difference (SEM)	NA	-0.7 (0.34)	-0.7 (0.32)
	95% CI	NA	(-1.4, -0.04)	(-1.4, -0.07)
	p-value	NA	0.038	0.030

Source: Table 8.2.2. (pg. 53) Clinical Study Report D1050106; Reviewer's MMRM Results
Remark: SEM stands for Standard Error of the Mean

Study D1050196

Primary Efficacy Analysis for the Primary Endpoint

This reviewer confirmed sponsor's primary efficacy analysis. The primary efficacy variable was the change from baseline to Week 6 in BPRS total score. Table 12 presents the ANCOVA results for this primary endpoint, using the LOCF approach. The change from baseline to Week 6 in the treatment group was statistically significantly greater than that in the placebo group (LS Mean Difference = –4.68, p = .0118).

Table 12. Study D1050196 BPRS Total Score LS Mean Change from Baseline to Week 6 (LOCF ANCOVA)

		Placebo	**SM-13496 80 mg**
No patients	N=180	90	90
Baseline	Mean (SD)	56.1 (6.84)	55.1 (5.95)
Change from Baseline	LS Mean (SEM)	–4.2 (1.36)	–8.9 (1.32)
Placebo-adjusted difference	LS Mean Difference	NA	–4.68
	95% CI	NA	(–8.3, –1.1)
	P-value	NA	0.0118

Source: Table 11.1.1. (pg. 57) and End-of-text Table 7.1 Clinical Study Report D1050196
Remark: SEM stands for Standard Error of the Mean

As seen from Table 13 , the observed treatment difference was numerically in favor of SM-13496 at every visit during the double blind phase.

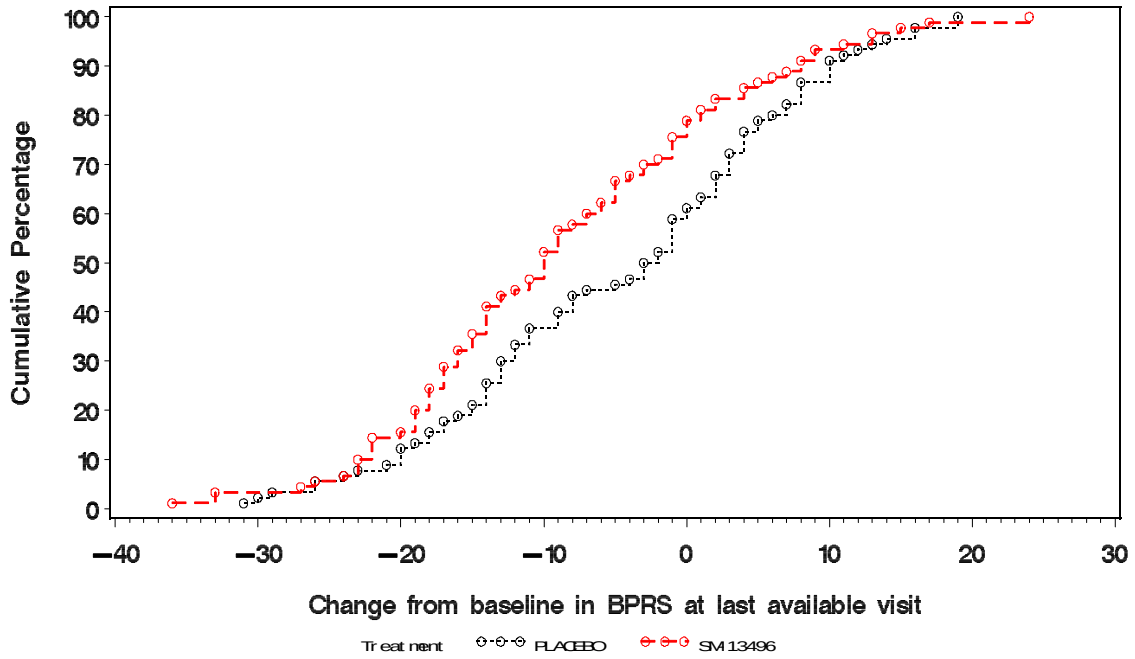
Table 13. Study D1050196 BPRS Total Score Mean Change from Baseline by Visit (LOCF ANCOVA).

	Placebo	SM-13496 80 mg	Treatment Difference: SM-13496 - Placebo	
Day	**LS Mean (SE)**	**LS Mean (SE)**	**LS Mean**	**95% CI**
3	-1.5 (0.60)	-3.7 (0.58)	-2.23	(-3.8, -0.6)
7	-2.2 (0.83)	-5.1 (0.81)	-2.86	(-5.1, -0.6)
14	-3.4 (0.99)	-6.6 (0.97)	-3.14	(-5.8, -0.5)
21	-4.0 (1.07)	-8.4 (1.05)	-4.42	(-7.3, -1.6)
28	-4.9 (1.23)	-8.7 (1.2)	-3.81	(-7.1, -0.5)
35	-3.8 (1.26)	-8.8 (1.23)	-5.0	(-8.4, -1.6)

Source: End-of-text Table 7.3 Clinical Study Report D1050196
Note: The reported 95% CIs are nominal CIs and are not adjusted for multiplicity.

Figure 3 displays empirical cumulative distribution functions (CDF) of the primary endpoint, change from baseline in BPRS score at week 6 (LOCF), for the two treatment arms in Study D1050196. Negative values of the primary endpoint represent improvement. The cumulative distribution functions describe the percentage of patients (vertical axis) in each treatment arm with primary endpoint values (horizontal axis) equal to or less than a given number x where x varies from -40 to 30. Approximately, for values of x less than 10, the CDF for SM-13496 80 mg treatment arm separates from the CDF for placebo arm. Numerically, in SM-13496 80 mg arm larger proportion of patients had negative value of the primary endpoint compared with the placebo arm. Because of the very high dropout rate, the graph should be interpreted with extra caution.

Figure 3. Study D1050196 Empirical cumulative distribution function



Source: Reviewer's result

Sensitivity Analysis for the Primary Endpoint

This reviewer conducted sensitivity analysis on the primary endpoint. The change from baseline to Week 6 in BPRS score was analyzed by mixed model with repeated measures (MMRM). The model included study center, treatment group, visit, and treatment group-by-visit interaction as factors and baseline BPRS total score as a covariate. The findings supported the primary analysis results. The SM-13496 at dose 80 mg was statistically significantly better than placebo.

Table 14. Study D1050196 LS Mean Change from Baseline in BPRS Score (MMRM Analysis)

	Placebo		**SM-13496 80 mg**		**Treatment Difference: SM-13496 - Placebo**	
Day	No	LS Mean (SE)	No	LS Mean (SE)	LS Mean	p-value
3	88	-1.5 (0.59)	89	-3.7 (0.58)	-2.2	0.0064
7	87	-2.5 (0.82)	85	-5.3 (0.81)	-2.8	0.0162
14	82	-4.2 (0.94)	73	-7.5 (0.96)	-3.3	0.0160
21	71	-5.1 (0.99)	63	-10.5 (1.02)	-5.4	0.0002
28	65	-6.4 (1.24)	59	-10.7 (1.28)	-4.3	0.0163
35	58	-4.9 (1.38)	53	-11.4 (1.44)	-6.4	0.0015
42	**49**	**-5.8 (1.55)**	**52**	**-11.3 (1.60)**	**-5.6**	**0.0131**

Source: Reviewer's Results

Supportive Secondary Endpoint (PANSS)

As an exploratory analysis, change from baseline in PANSS Total score was analyzed by ANCOVA model with terms for treatment, and pooled center, and PANSS baseline score as a covariate. SM-13496 80 mg treatment arm was statistically significantly better than placebo. This reviewer also considered MMRM analysis for the PANSS endpoint. Based on the MMRM analysis, SM-13496 80 mg was statistically significantly better than placebo.

Table 15. Study D1050196 PANSS Total Score LS Mean Change from Baseline to Week 6

		Placebo	**SM-13496 80 mg**
ANCOVA LOCF			
Change from Baseline	LS Mean (SEM)	-5.5 (2.17)	-14.1 (2.12)
Placebo-adjusted difference	LS Mean Difference	NA	-8.6 (2.94)
	P-value	NA	0.0040
MMRM			
Change from Baseline	LS Mean (SEM)	-7.8 (2.57)	-18.0 (2.67)
Placebo-adjusted difference	LS Mean Difference	NA	-10.2 (3.69)
	P-value	NA	0.0064

Source: Clinical Study Report D1050196

Secondary Endpoint

Table 16 presents sponsor's results for the change from baseline CGI-S score at Week 6, a secondary efficacy endpoint. At Week 6, statistically significant improvement was seen in the SM-13496 80 mg group relative to the placebo group in the LOCF ANCOVA (LS Mean Difference= –0.41, p = .0072) analysis with treatment group and study center as factors and baseline CGI-S score as covariate. This reviewer verified sponsor's results. Whether the magnitude of improvement was of clinical relevancy is deferred to the clinical review team.

This reviewer verified the sponsor's results and conducted exploratory analysis using MMRM. The SM-13496 80mg arm was significantly better than placebo arm (p-value of 0.0177).

Table 16. Study D1050196 Change from Baseline in CGI-S Score at Week 6 (ITT Population)

		Placebo	**SM-13496 80 mg**
No patients	N=180	90	90
Baseline	Mean (SD)	4.8 (0.67)	4.8 (0.71)
ANCOVA			
Change from Baseline	LS Mean (SEM)	-0.2 (0.11)	-0.6 (0.11)
Placebo-adjusted difference	LS Mean Difference	NA	-0.4 (0.15)
	95% CI	NA	(-0.7, -0.1)
	P-value	NA	0.0072
MMRM			
Change from Baseline	LS Mean (SEM)	-0.3 (0.14)	-0.8 (0.14)
Placebo-adjusted difference	LS Mean Difference	NA	-0.5 (0.20)
	95% CI	NA	(-0.9, -0.1)
	P-value	NA	0.0177

Source: Table 11.2.7.1 (pg. 61) and End-of-text Table 13.1 Clinical Study Report D1050196; Reviewer's results. Remark: SEM stands for Standard Error of the Mean

3.1.1.6 Reviewer's Comments.

In Study D1050006, the sponsor investigated efficacy of SM-13496 40 mg and SM-13496 120 mg. Study D1050006 had approximately 70% dropout rate. SM-13496 (Lurasidone) treatment groups were statistically superior to placebo in mean change from baseline to Week 6 in BPRS Total score (primary endpoint) whether based on LOCF ANCOVA (primary) or MMRM (sensitivity) analysis. The study results did not suggest additional benefit of 120 mg over the 40 mg based on the observed LS means differences. In addition, the strength of evidence may be weakened by the considerably high dropout rate along with the relatively small sample size in each group. The efficacy results in Study D1050006 should be interpreted with extra caution.

Study D1050196 investigated 80 mg of SM-13496. The study had nearly 50% dropout rate. Whether based on LOCF ANCOVA (primary) or MMRM (sensitivity) analysis, SM-13496 80 mg was statistically significantly better than placebo in mean change from baseline to Week 6 in BPRS Total score. The strength of evidence may be weakened by the considerably high dropout rate. Extra caution should be made for any interpretation of the results.

In both studies, all SM-13496 dose groups suggested statistically significant difference from placebo in change from baseline to Week 6 in CGI-S score, it was not pre-specified as a key secondary endpoint. Particularly in Study D1050006, there was no multiple testing procedure prespecified in the study protocol to adjust for both endpoints (BRS and CGI-S) simultaneously. Nevertheless, the p-values from both doses (40 mg and 120 mg) were very small so that any multiple testing procedure would lead to a statistical significance for both doses. Whether the magnitude of improvement in CGI-S was of clinical relevancy is deferred to the clinical review team.

3.1.2 STUDY D1050229 AND D1050231

3.1.2.1 Objective

The objective of studies D1050229 and D1050231 was to evaluate the efficacy of lurasidone compared with placebo in the treatment of subjects with acute schizophrenia (diagnosed by Diagnostic and Statistical Manual of Mental Disorders, 4th Ed. [DSM-IV™] criteria) as measured by the mean change from Baseline in the Positive and Negative Syndrome Scale (PANSS) total score at Week 6. Study D1050229 investigated efficacy of lurasidone at doses 40 mg/day, 80 mg/day, and 120 mg/day; Study D1050231 investigated efficacy of lurasidone at doses 40 mg/day and 120 mg/day.

3.1.2.2 Study Design

Both studies were 6-week, randomized, double-blind, multicenter, parallel-group studies designed to evaluate the safety and efficacy of lurasidone. Study D1050229 investigated three fixed doses of lurasidone (40 mg/day, 80 mg/day, or 120 mg/day) compared with placebo (1:1:1:1 treatment ratio). Study D1050231 had two lurasidone treatment arms (40 mg, and 120 mg), an active comparator arm (olanzapine 15 mg), and the placebo arm (1:1:1:1 treatment ratio).

3.1.2.3 Patient Disposition, Demographic and Baseline Characteristic

Study D1050229

Subject disposition is summarized in Table 17. Of the 500 subjects who were randomized to receive study medication, 328 subjects (66%) completed the double-blind phase of the study. Four subjects who were randomized received no study medication, one subject in the 40 mg group, 2 subjects in the 80 mg group, and one subject in the placebo group. The most common reason for early discontinuation was "insufficient clinical response or worsening of clinical condition" (81 patients; 16%).

Table 17. Study D1050229 Patient Population and Disposition

Patients	**Lurasidone 40 mg**	**Lurasidone 80 mg**	**Lurasidone 120 mg**	**Placebo**
Randomized	125 (100%)	123 (100%)	124 (100%)	128 (100%)
Received Study Drug	124 (99%)	121 (98%)	124 (100%)	127 (99%)
ITT Population	122 (98%)	119 (97%)	124 (100%)	124 (97%)
Completed DB Phase	84 (67%)	86 (70%)	85 (69%)	73 (57%)
Discontinued DB Phase	41 (33%)	37 (30%)	39 (31%)	55 (43%)
Insufficient clinical response	20 (16%)	7 (6%)	18 (15%)	32 (25%)
Adverse Event	6 (5%)	8 (7%)	7 (6%)	3 (2%)
Lost to follow-up	4 (3%)	2 (2%)	0 (0%)	6 (5%)
Protocol violation	0	0	0	0
Withdrew consent	9 (7%)	18 (15%)	12 (10%)	13 (10%)
Administrative				

Source: Corresponds to Table 11 (p.67)

Demographic data are summarized in Table 18 (Safety population). Of the 496 subjects in the Safety population, 346 (70%) were male and 150 (30%) were female. Subject age ranged from 18 to 72 years, with a mean age of 39.0 years. The largest racial subgroup was White (49%), followed by Black or African American (34%), and Asian (15%). Native Americans and Native Hawaiian or Other Pacific Islanders made up less than 1% each of the Safety population. Demographics of the ITT population were similar.

Table 18. D1050229 Demographic and Baseline Characteristics (Safety Population)

Variable	**Lurasidone 40 mg N=124**	**Lurasidone 80 mg N=121**	**Lurasidone 120 mg N=124**	**Placebo N=127**
Male	83 (67%)	78 (64%)	92 (74%)	93 (73%)
Female	41 (33%)	43 (36%)	32 (26%)	34 (27%)
Age				
Mean (min, max)	40.7 (18, 72)	38.7 (19, 62)	37.7 (18, 65)	38.1 (20, 64)
Race				
Caucasian	57 (46%)	61 (50%)	60 (48%0	66 (52%)
Black	49 (40%)	40 (33%)	41 (33%)	38 (30%)
Asian	17 (14%)	19 (16%)	20 (16%)	20 (16%)
Other	1 (<1%)	1 (<1%)	3 (2%)	3 (2%)
Baseline PANSS Total Score (ITT)	**N=122**	**N=119**	**N=124**	**N=124**
Mean (SD)	96.5 (11.5)	96.0 (10.8)	96.0 (9.7)	96.8 (11.1)

Source: Table 14 (p.72), Table 15 (p.74) Clinical Study Report D1050229

Study D1050231

Subject disposition is summarized in Table 19. Of the 478 subjects who were randomized to receive study medication, 298 subjects (62%) completed the double-blind phase of the study. Three subjects who were randomized received no study medication, one subject in the lurasidone 40 mg group (Subject 23102806), 1 subject in the lurasidone 120 mg group (Subject 23105216), and one subject in the olanzapine group (Subject 23114104). The most common reason for early discontinuation from the double-blind phase of the study was withdrawal of consent, with 16 subjects (13%) in the lurasidone 40 mg group, 28 subjects (24%) in the lurasidone 120 mg group, 19 subjects (15%) in the olanzapine group, and 12 subjects (10%) in the placebo group discontinuing due to withdrawal of consent. Overall, 61 subjects (13%) discontinued from the double-blind phase of the study for "insufficient clinical response or worsening of existing condition".

Table 19. Study D1050231 Patient Population and Disposition

Patients	Lurasidone 40 mg	Lurasidone 120 mg	Olanzapine 15 mg	Placebo
Randomized	120 (100%)	119 (100%)	123 (100%)	116 (100%)
Received Study Drug	119 (99%)	118 (99%)	122 (99%)	116 (100%)
ITT Population	119 (99%)	118 (99%)	122 (99%)	114 (98%)
Completed DB Phase	77 (64%)	66 (55%)	84 (68%)	71(61%)
Discontinued DB	43 (36%)	53 (45%)	39 (32%)	45 (39%)
Insufficient clinical response	16 (13%)	9 (8%)	8 (7%)	18 (16%)
Adverse Event	8 (7%)	14 (12%)	8 (7%)	10 (9 %)
Lost to follow-up	1(<1%)	2 (2%)	1 (<1%)	2 (2%)
Protocol violation	2 (2%)	0 (0%)	0 (0%)	1 (<1%)
Withdrew consent	16 (13%)	28 (24%)	19 (15%)	12 (10%)
Administrative	0 (0%)	0 (0%)	3 (2%)	2 (2%)

Source: Corresponds to Table 11 (p. 69)

Demographic data are summarized in Table 20 (ITT population). Of the 473 subjects in the ITT population, 369 (78%) were male and 104 (22%) were female. Subject age ranged from 18 to 68 years, with a mean age of 37.7 years. The largest racial subgroup was White (36%), followed by Black or African American (34%), and Asian (24%).

Table 20. D1050231 Demographic and Baseline Characteristics (ITT Population)

Variable	Lurasidone 40 mg N=119	Lurasidone 120 mg N=118	Olanzapine 15 mg N=122	Placebo N=114
Gender				
Male	93 (78%)	93 (79%)	95 (78%)	88 (77%)
Female	26 (22%)	25 (21%)	27 (22%)	26 (23%)
Age				
Mean (min, max)	37.7 (18, 63)	37.9 (18, 68)	38.3 (19, 62)	37.0 (18, 64)
Race				
Caucasian	44 (37%)	48 (41%)	41 (34%)	36 (32%)
Black	39 (33%)	36 (31%)	44 (36%)	41 (36%)
Asian	31 (26%)	27 (23%)	30 (25%)	28 (24%)
Other	5 (4%)	7 (5%)	7 (5%)	10 (8%)
Baseline PANSS Total Score				
Mean (SD)	96.6 (10.7)	97.9 (11.3)	96.3 (12.2)	95.8 (10.8)

Source: Table 19 (p. 85), Table 14.1.3.1 Clinical Study Report D1050231

3.1.2.4 Statistical Methodologies and Endpoints

For both studies, the primary population for the efficacy analysis was the Intent-to-Treat (ITT) population. All subjects who were randomized, received at least one dose of study medication, and have a baseline efficacy measurement and at least one post-Baseline efficacy measurement, were included in the efficacy analysis in the treatment group to which they were randomized.

Primary Analysis for the Primary Endpoint

The primary analysis for the primary efficacy endpoint, the change from Baseline PANSS total score at Week 6, was based on MMRM model under the assumption of an unstructured covariance matrix (Liang-2000). The model included factors for pooled center, time (including all scheduled post-Baseline visits, modeled as a categorical variable), Baseline PANSS total score, treatment, and treatment-by-time interaction. The Kenward-Rogers method was used to estimate the denominator degrees of freedom. Treatment differences were evaluated via contrasts for the time-by-treatment factor.

Supportive Analysis for the Primary Endpoint

As a supportive analysis, the change from Baseline in PANSS total scores at Week 6 was evaluated using LOCF ANCOVA, with effects for Baseline total PANSS score, pooled center, and treatment.

Key Secondary Endpoint

The change from Baseline in CGI-S at Week 6 was evaluated using the same MMRM model used for the PANSS total score, and was included in the Hommel-based tree-gatekeeping procedure to adjust for multiple comparisons.

Study Center Pooling

All centers with 7 or fewer subjects were pooled. Small centers were pooled by size within country or geographic region if necessary, with the intention that no pooled center would contain more than 24 randomized subjects. Centers with 8 or more randomized subjects were not pooled.

Multiplicity Adjustment for both the Primary and Key Secondary Endpoints

Study D1050229:

The Hommel-based tree-gatekeeping procedure was applied to p-values from the MMRM analysis to control the family-wise Type I error rate at 5% by taking into account multiple doses and multiple primary and key secondary endpoints. The hypotheses associated with the primary and key secondary variables for efficacy claim were grouped into 2 hierarchical families:

Family F1: lurasidone 40 mg/day versus placebo (H1), lurasidone 80 mg/day versus placebo (H2), and lurasidone 120 mg/day versus placebo (H3) based on change from Baseline in PANSS total score at Week 6 (E1);

Family F2: lurasidone 40 mg/day versus placebo (H4), lurasidone 80 mg/day versus placebo (H5), and lurasidone 120 mg/day versus placebo (H6) based on change from Baseline in CGI-S at Week 6 (E2);

The gatekeeping procedure accounted for the logical restrictions in this problem by performing multiplicity adjustment in two steps:

Step 1: The lurasidone-placebo comparisons for E1 (hypotheses H1, H2, and H3) were performed using a truncated version of the Hommel test.

Step 2: The lurasidone-placebo comparisons for E2 (hypotheses H4, H5, and H6), corresponding to the doses that were significant at Step 1, were performed using a truncated version of the Hommel test. For example, H4 was tested only if H1 was rejected.

The value of the truncation parameter Gamma 1 used to determine the balance of power in Families 1, and 2 was set at Gamma1=0.5. The Hommel-based tree-gatekeeping procedure controlled the overall Type I error rate in the strong sense at the α level.

Study D1050231:
The Hommel-based tree-gatekeeping procedure (described in Appendix 4 of the SAP) was applied to p-values from the mixed models for repeated measurements (MMRM) analysis to control the family-wise Type I error rate at 5% by taking into account multiple doses and multiple primary and key secondary endpoints. The hypotheses associated with the primary and key secondary variables for efficacy claim were grouped into 2 hierarchical families:
Family F1: lurasidone 40 mg/day versus placebo (H1) and lurasidone 120 mg/day versus placebo (H2) based on change from Baseline in PANSS total score at Week 6 (E1);
Family F2: lurasidone 40 mg/day versus placebo (H3) and lurasidone 120 mg/day versus placebo (H4) based on change from Baseline in CGI-S at Week 6 (E2);
The gatekeeping procedure accounted for the logical restrictions in this problem by performing multiplicity adjustment in 2 steps:
Step 1: The lurasidone-placebo comparisons for E1 (hypotheses H1 and H2) were performed using a truncated version of the Hommel test.
Step 2: The lurasidone-placebo comparisons for E2 (hypotheses H3 and H4), corresponding to the doses that were significant at Step 1, were performed using a regular Hommel test. For example, H4 was tested only if H2 was rejected.
The value of the truncation parameter Gamma1 used to determine the balance of power in Families 1 and 2 was set at Gamma1=0.0. The Hommel-based tree-gatekeeping procedure controlled the overall Type I error rate in the strong sense at the α level.

3.1.2.5 Results of Efficacy Analysis

Study D1050229

Primary Analysis for the Primary Endpoint
The primary analysis for the primary efficacy endpoint, the change from Baseline in PANSS total score at Week 6, was based on MMRM model under the assumption of an unstructured covariance matrix. The model included factors for pooled center, time (including all scheduled post-Baseline visits, modeled as a categorical variable), baseline PANSS total score, treatment, and treatment-by-time interaction. The PANSS total score LS mean change from Baseline (repeated measures) is summarized in Table 21.

After adjustment for multiplicity using the Hommel-based tree-gatekeeping procedure with the truncation parameter Gamma=0.5, there was a statistically significant treatment difference with placebo of -6.4 (p=0.034) for the lurasidone 80 mg treatment arm. Neither the 40 mg group nor the 120 mg group separated from placebo at Week 6. This reviewer confirmed the sponsor's results.

Table 21. Study D1050229 PANSS Total Score LS Mean Change from Baseline to Week 6 (MMRM)

Week/ Day	**Lurasidone 40 mg vs. Placebo**		**Lurasidone 80 mg vs. Placebo**		**Lurasidone 120 mg vs. Placebo**		**Placebo**
	LS Mean Difference	p-value	LS Mean Difference	p-value	LS Mean Difference	p-value	LS Mean
Day 4	-1.1 (0.9)	0.241	-1.0 (0.9)	0.296	-1.0 (0.9)	0.263	-3.2 (0.7)
Week 1	-1.0 (1.2)	0.433	-2.0 (1.3)	0.113	-1.9 (1.2)	0.122	-6.3 (0.9)
Week 2	-1.5 (1.6)	0.360	-3.5 (1.6)	0.031	-3.4 (1.6)	0.036	-9.4 (1.1)
Week 3	-2.7 (1.9)	0.156	-4.6 (1.9)	0.018	-4.3 (1.9)	0.026	-11.8 (1.3)
Week 4	-2.2 (2.1)	0.301	-5.1 (2.1)	0.017	-3.9 (2.1)	0.062	-14.1 (1.5)
Week 5	-2.3 (2.3)	0.304	-5.9 (2.3)	0.010	-4.2 (2.3)	0.064	-15.3 (1.6)
Week 6	**-2.1 (2.5)**	**0.394**	**-6.4 (2.5)**	**0.011**	**-3.5 (2.5)**	**0.163**	-17.0 (1.8)
95% CI	**(-7.0, 2.8)**		**(-11.3, -1.5)**		**(-8.4, 1.4)**		
Adjusted p-value	**0.591**		**0.034**		**0.391**		

Source: Table 14.2.1.1. Clinical Study Report D1050229

Supportive Analysis for the Primary Endpoint

The sponsor also conducted supportive analysis for the primary endpoint based on LOCF ANCOVA with treatment and pooled center as fixed factors and baseline value as a covariate (see Table 22). The supportive analysis confirmed primary efficacy results. Lurasidone 80 mg was statistically superior to Placebo. Neither the 40 mg group nor the 120 mg group separated from placebo at Week 6.

Table 22. Study D1050229 PANSS Total Score LS Mean Change from Baseline to Week 6 (LOCF ANCOVA)

	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Placebo
No patients	121	118	123	124
LS Mean change from baseline (SE)	-17.4 (1.6)	-20.8 (1.6)	-18.5 (1.6)	-14.7 (1.6)
Placebo-adjusted Difference	-2.7 (2.2)	-6.1 (2.3)	-3.8 (2.2)	NA
p-value	0.236 (0.354 Hommel-based)	0.007 (0.021 Hommel-based)	0.086 (0.206 Hommel-based)	NA

Source: Table 19 Clinical Study Report D1050229 (p. 84)

Key Secondary Endpoint

The key secondary efficacy endpoint was the change from baseline in CGI-S score at Week 6. At the end of six weeks of double-blind treatment, patients receiving Lurasidone 80 mg showed statistically significantly greater improvement relative to placebo-treated patients in the CGI-S score using the MMRM approach. Whether the magnitude of improvement was of clinical relevancy is deferred to the clinical review team. Lurasidone 40 mg and Lurasidone 120 mg did not demonstrate superiority versus placebo. These results were confirmed by the reviewer.

Table 23 Study D1050229 CGI-S Score LS Mean Change from Baseline to Week 6 (MMRM)

Week/Day	**Lurasidone 40 mg vs. Placebo**		**Lurasidone 80 mg vs. Placebo**		**Lurasidone 120 mg vs. Placebo**		**Placebo**
	LS Mean Diff. (SE)	p-value	LS Mean Diff. (SE)	p-value	LS Mean Diff. (SE)	p-value	LS Mean
Day 4	-0.0 (0.1)	0.478	-0.0 (0.1)	0.404	-0.0 (0.1)	0.734	-0.1 (0.0)
Week 1	-0.0 (0.1)	0.516	-0.1 (0.1)	0.156	-0.1 (0.1)	0.366	-0.2 (0.1)
Week 2	-0.0 (0.1)	0.784	-0.2 (0.1)	0.025	-0.1 (0.1)	0.429	-0.5 (0.1)
Week 3	-0.1 (0.1)	0.378	-0.3 (0.1)	0.029	-0.2 (0.1)	0.178	-0.7 (0.1)
Week 4	-0.1 (0.1)	0.266	-0.3 (0.1)	0.009	-0.2 (0.1)	0.048	-0.8 (0.1)
Week 5	-0.1 (0.1)	0.458	-0.4 (0.1)	0.006	-0.3 (0.1)	0.029	-0.9 (0.1)
Week 6	**-0.1 (0.1)**	**0.365**	**-0.4 (0.1)**	**0.005**	**-0.2 (0.1)**	**0.169**	**-1.0 (0.1)**
95 % CI	**(-0.4, 0.1)**		**(-0.7, -0.1)**		**(-0.5, 0.1)**		
Adjusted p-value*	**0.591**		**0.034**		**0.543**		

Source: Table 14.2.2.1. Clinical Study Report D1050229 (* Adjusted p-values refer to multiple-dose adjustment at week 6, not across all visits.)

Study D1050231

Primary Analysis for the Primary Endpoint

The primary analysis for the primary efficacy endpoint, the change from Baseline in PANSS total score at Week 6, was based on MMRM model under the assumption of an unstructured covariance matrix. The model included factors for treatment, pooled center, time (including all scheduled post-Baseline visits, modeled as a categorical variable), baseline PANSS total score, and treatment-by-time interaction. The LS mean change from Baseline (repeated measures) in PANSS total score is summarized in Table 24. After multiplicity adjustment using the Hommel-based tree-gatekeeping procedure, there were statistically significant treatment differences with placebo at Week 6 for lurasodone 40 mg (-9.7, p=0.002) and lurasodone 120 mg (-7.5, p=0.022). This reviewer confirmed the sponsor's results. However, the 120 mg dose did not seem to add additional benefit to the 40 mg dose.

Table 24. Study D1050231 PANSS Total Score LS Mean Change from Baseline to Week 6 (MMRM)

Week/ Day	**Lurasidone 40 mg vs. Placebo**		**Lurasidone 120 mg vs. Placebo**		**Olanzapine 15 mg vs. Placebo**		**Placebo**
	LS Mean Diff	p-value	LS Mean Diff	p-value	LS Mean Diff	p-value	LS Mean
Day 4	-0.2 (1.0)	0.798	-0.6 (1.0)	0.559	-1.3 (0.9)	0.166	-4.8 (0.7)
Week 1	-3.1 (1.3)	0.022	-1.7 (1.3)	0.201	-3.5 (1.3)	0.008	-7.0 (1.0)
Week 2	-4.6 (1.7)	0.008	-3.2 (1.8)	0.073	-5.4 (1.7)	0.002	-10.4 (1.2)
Week 3	-7.0 (2.2)	0.002	-6.5(2.2)	0.004	-9.5 (2.2)	<0.001	-11.4 (1.6)
Week 4	-8.1 (2.4)	<0.001	-8.2 (2.5)	<0.001	-11.4 (2.4)	<0.001	-13.1 (1.7)
Week 5	-8.9 (2.7)	0.001	-9.6 (2.8)	<0.001	-11.9 (2.7)	<0.001	-15.0 (1.9)
Week 6	**-9.7 (2.9)**	**<0.001**	**-7.5 (3.0)**	**0.011**	**-12.6 (2.8)**	**<0.001**	-16.0 (2.1)
95 % CI	**(-15.3,-4.1)**		**(-13.4,-1.7)**		**(-18.2,-7.9)**		
Adjusted p-value	**0.002**		**0.022**		**Unadjusted p-value <0.001**		

Source: Table 14.2.1.1. Clinical Study Report D1050231

Supportive Analysis for the Primary Endpoint

The sponsor also conducted supportive analysis for the primary endpoint based on LOCF ANCOVA with treatment and pooled center as fixed factors and baseline value as a covariate (see Table 25). The supportive analysis results are not fully consistent with the primary analysis results. Although lurasidone 40 mg was statistically superior to Placebo, the lurasidone 120 mg treatment arm did not separate from placebo at Week 6. The result also suggests that the 120 mg did not seem to add additional benefit over the 40 mg dose.

Table 25. Study D1050231 PANSS Total score LS Mean Change from Baseline to Week 6 (LOCF ANCOVA)

	Lurasidone 40 mg	Lurasidone 120 mg	Olanzapine 15 mg	Placebo
No patients	118	118	121	114
LS Mean change from baseline (SE)	-23.1 (1.7)	-20.0 (1.7)	-26.7 (1.7)	-15.2 (1.7)
Placebo-adjusted Difference	-7.9 (2.4)	-4.8 (2.4)	-11.4 (2.4)	NA
p-value	0.001	0.049	<0.001	NA
Adjusted p-value	0.001 (Hommel)	0.098 (Hommel)		

Source: Table 19 Clinical Study Report D1050231 (p. 86)

Key Secondary Endpoint

The key secondary endpoint was the change from baseline in CGI-S score at Week 6. At the end of six weeks of double-blind treatment, patients receiving lurasidone 40 mg and lurasidone 120 mg showed statistically significantly greater improvement relative to placebo-treated patients in the CGI-S score using the MMRM approach. This reviewer confirmed sponsor's results. However, whether the magnitude of improvement is of clinical relevancy is deferred to the clinical review team.

Table 26. Study D1050231 CGI-S Score LS Mean Change from Baseline to Week 6 (MMRM)

Week/Day	**Lurasidone 40 mg vs Placebo**		**Lurasidone 120 mg vs Placebo**		**Olanzapine 15 mg vs Placebo**		**Placebo**
	LS Mean Difference	p-value	LS Mean Difference	p-value	LS Mean Difference	p-value	LS Mean
Day 4	0.0 (0.1)	0.540	0.0 (0.1)	0.920	-0.0 (0.1)	0.473	-0.2 (0.0)
Week 1	-0.1 (0.1)	0.259	-0.2 (0.1)	0.038	-0.1 (0.1)	0.096	-0.3 (0.1)
Week 2	-0.2 (0.1)	0.012	-0.2 (0.1)	0.015	-0.3 (0.1)	0.006	-0.5 (0.1)
Week 3	-0.3 (0.1)	0.022	-0.3 (0.1)	0.008	-0.4 (0.1)	<0.001	-0.7 (0.1)
Week 4	-0.3 (0.1)	0.014	-0.4 (0.1)	0.003	-0.5 (0.1)	<0.001	-0.8 (0.1)
Week 5	-0.4 (0.1)	0.004	-0.5 (0.1)	<0.001	-0.5 (0.1)	0.001	-0.9 (0.1)
Week 6	**-0.4 (0.1)**	**0.006**	**-0.3 (0.1)**	**0.040**	**-0.5 (0.1)**	**<0.001**	**-1.1 (0.1)**
	95 % CI	**(-0.7, -0.1)**	**95% CI**	**(-0.6, -0.0)**	**95% CI**	**(-0.8, -0.2)**	
	Adjusted p-value	**0.011**	**Adjusted p-value**	**0.040**	**Adjusted p-value**	**<0.001**	

Source: Table 14.2.2.1. Clinical Study Report D1050231

3.1.2.6 Reviewer's Comments

For both studies, D1050229 and D1050231, the primary efficacy variable was change from baseline to Week 6 in PANSS total score and the key secondary variable was change from baseline to Week 6 in CGI-S score. For both endpoints, the primary efficacy analysis model was MMRM.

In study D1050229, there were three Lurasidone arms (40 mg, 80 mg, and 120 mg) and one placebo arm. The sponsor used Hommel-based gatekeeping procedure to control studywise type I error rate. The hypotheses considered for inclusion in labeling were divided into two families. The primary family included null hypotheses associated with the primary endpoint: comparison of three doses of Lurasidone with Placebo in change from Baseline in PANSS total score. The secondary family consisted of null hypotheses associated with the key secondary endpoint, change from baseline in CGI-S. For the Hommel-based testing procedure, the sponsor pre-specified parameter gamma as 0.5. The Hommel-based gatekeeping procedure is a closure-based multiple testing procedure where for each intersection, null hypotheses from the primary family are tested by truncated Hommel procedure with truncation parameter gamma, and null hypotheses from the secondary family are tested by the Hommel procedure. After adjusting for multiplicity, Lurasidone 80 mg treatment arm was statistically significantly better than placebo in both, the primary and the key secondary, endpoints with adjusted p-values 0.034 and 0.034 respectively.

Study D1050231 included four treatment arms: two lurasidone treatment arms (40 mg and 120 mg), placebo, and active comparator (Olanzapine 15 mg). The same type of multiple adjustment procedure, the Hommel-based gatekeeping procedure, was applied to compare two doses of Lurasidone with placebo. The primary family included null hypotheses related to the primary endpoint, and the secondary family included null hypotheses associated with the key secondary endpoint. The value of truncation parameter gamma was set to be equal to zero. Numerically, olanzapine 15 mg showed better treatment effect compared to lurasidone and placebo arms. After multiplicity adjustment, both lurasidone treatment arms were statistically significantly superior to placebo in both the primary and the key secondary endpoints. However, the 120 mg dose did not seem to add additional benefit over the 40 mg dose.

For both studies, whether the magnitude of improvement in CGI-S was of clinical relevancy is deferred to the clinical review team.

3.2 EVALUATION OF SAFETY

Not evaluated by this reviewer.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 GENDER, RACE AND AGE

4.1.1 STUDIES D1050106 AND D1050196

This reviewer conducted exploratory subgroup analysis on the primary efficacy variable (change from baseline in BPRS Total score at Week 6), using LOCF ANCOVA models, including terms

for treatment and the baseline score. The subgroups of interest included gender and race. For all subgroups, and for both studies, the treatment effect appeared to be numerically in favor of SM-13496 when compared with placebo.

Table 27. Study D1050006 Subgroup Analysis: BPRS Total Score LS Mean Change from Baseline (LOCF ANCOVA).

Subgroup	Placebo	SM-13496 40 mg	SM-13496 120 mg
Gender			
Male N=111	-4.2 (1.66) N=42	-8.9 (1.81) N=35	-10.3 (1.83) N=34
Female N=34	-0.5 (3.72) N=7	-11.2 (2.68) N=14	-10.3 (2.77) N=13
Race			
White N=59	-5.1 (2.43) N=19	-7.5 (2.38) N=20	-8.5 (2.39) N=20
Black N=73	-3.0 (2.18) N=25	-9.7 (2.23) N=24	-11.9 (2.23) N=24
Other N=13	-3.5 (3.73) N=5	-12.7 (3.71) N=5	-12.3 (4.58) N=3

Source: Reviewer's Results

Table 28. Study D1050196 Subgroup Analysis: BPRS Total score LS Mean Change from Baseline (LOCF ANCOVA).

Subgroup	Placebo	SM-13496 80 mg	Treatment Difference
Gender			
Male N =138	-4.8 (1.42) N=70	-8.7 (1.44) N=68	-3.9 (2.02)
Female N=42	-3.2 (2.81) N=20	-9.7 (2.7) N=22	-6.5 (3.89)
Race			
White N=61	-4.3 (2.19) N=26	-8.4 (1.89) N=35	-4.1 (2.89)
Black N= 103	-4.1 (1.69) N=56	-8.9 (1.84) N=47	-4.8 (2.50)
Other N=13 N=26	-7.9 (4.33) N=8	-9.9 (4.33) N=8	-2.0 (6.27)

Source: Reviewer's Results

4.1.2 STUDIES D1050229 AND D1050231.

This reviewer conducted exploratory subgroup analysis on the primary efficacy variable (change from baseline in PANSS Total score at week 6), using MMRM models, including the terms for treatment, visit and treatment by visit interaction, and the baseline score as a covariate. The subgroups of interest included gender and race. For all subgroups, and for both studies, except African American racial subgroup for the 120 mg treatment arm in Study D1050229, the treatment effect appeared to be numerically in favor of lurasidone when compared with placebo.

Table 29. Study D1050229 Subgroup Analysis: PANSS Total Score LS Mean Change from Baseline (MMRM).

Subgroup	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Placebo
Gender				
Male N=337	-17.5 (2.3) N=81	-23.7 (2.3) N=75	-18.5 (2.2) N=91	-16.8 (2.2) N=90
Female N=149	-23.4 (2.7) N=40	-23.1 (2.7) N=43	-26.4 (3.1) N=32	-18.1 (3.2) N=34
Race				
White N=240	-18.0 (2.7) N=55	-24.6 (2.6) N=59	-18.1 (2.7) N=60	-14.6 (2.5) N=66
Black or African American; N=164	-20.8 (2.7) N=49	-19.3 (3.0) N=39	-18.8 (2.9) N=40	-20.2 (3.1) N=36
Other N=82	-20.0 (4.7) N=17	-28.6 (4.5) N=20	-29.0 (4.1) N=23	-19.7 (4.4) N=22

Source: Reviewer's Results

Table 30. Study D1050231 Subgroup Analysis: PANSS Total Score LS Mean Change from Baseline (MMRM).

Subgroup	Lurasidone 40 mg	Lurasidone 120 mg	Olanzapine 15 mg	Placebo
Gender				
Male N=367	-25.4 (2.2) N=92	-21.3 (2.4) N=93	-29.2 (2.1) N=94	-15.2 (2.3) N=88
Female N=104	-26.5 (4.5) N=26	-31.5 (4.8) N=25	-26.2 (4.2) N=27	-17.6 (4.3) N=26
Race				
White N=167	-23.5 (3.2) N=43	-17.7 (3.3) N=48	-24.8 (3.1) N=40	-13.3 (3.5) N=36
Black or African American; N=160	-20.4 (2.8) N=39	-24.1 (2.9) N=36	-25.5 (2.6) N=44	-14.6 (2.7) N=41
Other N=144	-33.2 (3.9) N=36	-30.1 (4.5) N=34	-36.3 (4.0) N=37	-20.4 (4.2) N=37

Source: Reviewer's Results

4.2 OTHER SPECIAL/SUBGROUP POPULATIONS

4.2.1 STUDIES D1050229 AND D1050231

This reviewer conducted exploratory regional subgroup analysis (US, Other) on the primary efficacy variable (change from baseline in PANSS Total score at week 6), using MMRM models, including terms treatment, visit and treatment by visit interaction, and the baseline score as a covariate. In Study D1050229, the placebo arm was numerically superior to lurasidone 40 mg and 120 mg treatment arms, and it appears that the observed treatment effects (regardless of the doses) were mainly driven by patients from non-US. In study D1050231, for both regional subgroups the treatment effect appeared to be numerically in favor of lurasidone when compared with placebo.

Table 31. Study D1050229 Subgroup Analysis: PANSS Total Score LS mean Change from Baseline (MMRM).

Subgroup	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Placebo
Region				
United States N=268	-17.0 (2.4) N=69	-20.1 (2.4) N=63	-17.3 (2.4) N=69	-18.1 (2.4) N=67
Other N=218	-22.1 (2.6) N=52	-27.1 (2.5) N=55	-24.3 (2.6) N=54	-16.5 (2.6) N=57

Source: Reviewer's Results

Table 32. Study D1050231 Subgroup Analysis: PANSS Total Score LS Mean Change from Baseline (MMRM).

Subgroup	Lurasidone 40 mg	Lurasidone 120 mg	Olanzapine 15 mg	Placebo
Region				
United Sates N=281	-20.0 (2.3) N=69	-17.5 (2.4) N=72	-23.0 (2.1) N=73	-12.8 (2.3) N=67
Other N=190	-32.5 (3.1) N=49	-32.6 (3.5) N=46	-36.2 (3.2) N=48	-21.2 (3.4) N=47

Source: Reviewer's Results

5 SUMMARY AND CONCLUSIONS

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE

Studies D1050006 and D1050196

In Study D1050006, the sponsor investigated efficacy of SM-13496 40 mg and SM-13496 120 mg. Study D1050006 had approximately 70% dropout rate. Based on LOCF ANCOVA analysis, SM-13496 (Lurasidone) treatment groups were statistically superior to placebo in mean change from baseline to Week 6 in BPRS Total score (primary endpoint). The Dunnett's adjusted p-values of pairwise comparisons with placebo were 0.018 (SM-13496 40 mg versus Placebo) and 0.004 (SM-13496 120 mg versus Placebo). However, the 120 mg dose did not seem to add additional benefit over the 40 mg dose. In addition, the strength of evidence may be weakened by the considerably high dropout rate along with the relatively small sample size in each group. The efficacy results in Study D1050006 should be interpreted with extra caution.

Study D1050196 had two treatment arms: SM-13496 80 mg and placebo. The study had 50% dropout rate. Based on LOCF ANCOVA analysis, SM-13496 80 mg was statistically significantly better than placebo in mean change from baseline to Week 6 in BPRS Total score (p-value 0.0118). The strength of evidence may be weakened by the considerably high dropout rate. Extra caution should be made for any interpretation of the results.

In both studies, change from baseline in CGI-S score was a secondary endpoint, but not a pre-specified key secondary endpoint. No multiplicity adjustment for null hypotheses associated with CGI-S was prespecified in study D1050006. Nevertheless, in study D1050006 the LOCF ANCOVA p-values from both doses (40 mg and 120 mg) were very small so that any multiple testing procedure would lead to a statistical significance for both doses. Study D1050196 investigated 80 mg only, and the p-value from testing this endpoint was also very small. Whether

the magnitude of improvement in CGI-S was of clinical relevancy is deferred to the clinical review team.

Studies D1050229 and D1050231

In Study D1050229 three doses of lurasidone (40 mg, 80 mg, and 120 mg) were compared with placebo with respect to the primary endpoint (change from baseline in PANSS total score) and the key secondary endpoint (change from baseline in CGI-S). After multiplicity adjustment using the Hommel-based gatekeeping procedure with truncation parameter gamma=0.5, lurasidone 80 mg was statistically significantly better than placebo in both, the primary and the key secondary endpoints. Doses of 40 mg and 120 mg failed to demonstrate efficacy for either of the two endpoints.

Study D1050231 included four treatment arms: two lurasidone treatment arms (40 mg and 120 mg), placebo, and active comparator (Olanzapine 15 mg). Numerically, Olanzapine 15 mg showed better treatment effect compared to lurasidone and placebo arms. After multiplicity adjustment, using the Hommel-based gatekeeping procedure with truncation parameter zero, both lurasidone treatment arms (40 mg and 120 mg) were statistically significantly superior to placebo in both, the primary and the key secondary endpoints.

Whether the magnitudes of improvement in CGI-S in studies D1050229 and D1050231 were of clinical relevancy is deferred to the clinical review team.

Overall summary of the efficacy results for the primary endpoints

Change from baseline in BPRS: Lurasidone 40 mg and Lurasidone 120 mg demonstrated efficacy versus placebo in Study D1050106. Lurasidone 80 mg demonstrated efficacy versus placebo in Study D1050196.

Change from baseline in PANSS total score: Lurasidone 80 mg demonstrated efficacy in Study D1050229. Lurasidone at doses 40 mg and 120 mg was statistically significantly better than placebo in Study D1050231, but failed to demonstrate efficacy in Study D1050229.

Table 33. Summary of Efficacy Results for The Primary Endpoints

		Treatment Difference from Placebo			
Study	**Primary Endpoint**	**Lurasidone 40 mg**	**Lurasidone 80 mg**	**Lurasidone 120 mg**	**Olanzapine 15 mg**
D1050006	BPRS (ANCOVA)	-5.6 p=0.018 (Dunnett)	NA	-6.7 p=0.004 (Dunnett)	NA
D1050196	BPRS (ANCOVA)	NA	-4.7 p=0.0118	NA	NA
D1050229	PANSS (MMRM)	-2.1 p=0.591 (Hommel-based)	-6.4 p=0.034 (Hommel-based)	-3.5 p=0.391 (Hommel-based)	NA
D1050231	PANSS (MMRM)	-9.7 p=0.02 (Hommel-based)	NA	-7.5 p=0.022 (Hommel-based)	-12.6 p<0.001

Source: Clinical study reports D1050006, D1050196, D1050229, D1050231; p-values were adjusted using pre-specified multiple testing procedures: Dunnett procedure for Study D1050006; the Hommel-based gatekeeping procedure for studies D1050229, D1050231

Overall summary of the efficacy results for PANSS

Table 34 summarizes efficacy results for change from baseline in PANSS total score based on LOCF analysis. After adjustment for multiplicity, Lurasidone 40 mg was superior to placebo in Study D1050231 (1 out of 3 studies); Lurasidone 80 mg was superior to placebo in Studies D1050196 and Study D1050229 (2 out of 2 studies); Lurasidone 120 mg was superior to Placebo in Study D1050006 (1 out of 3 studies).

Table 34. Summary of Efficacy Results for Change from Baseline in PANSS Total Score Based on LOCF ANCOVA Analysis.

Study	Primary Endpoint	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Olanzapine 15 mg
D1050006	PANSS (ANCOVA)	-7.6 p=0.076 (Dunnett)	NA	-11 p=0.009 (Dunnett)	NA
D1050196	PANSS (ANCOVA)	NA	-8.6 p=0.0040	NA	NA
D1050229	PANSS (ANCOVA)	-2.7 p=0.354 (Hommel-based)	-6.1 p=0.021 (Hommel-based)	-3.8 p=0.206 (Hommel-based)	NA
D1050231	PANSS (ANCOVA)	-7.9 p=0.002 (Hommel-based)	NA	-4.8 p=0.098 (Hommel-based)	-11.4 p<0.001

Source: Clinical study reports D1050006, D1050196, D1050229, D1050231

Based on MMRM analysis, lurasidone at doses 40 mg and 120 mg was superior to placebo in Study D1050006 and Study D1050231, but failed to demonstrate efficacy in Study D1050229. Lurasidone 80 mg demonstrated efficacy in Study D1050196 and Study D1050229.

Table 35. Summary of Efficacy Results for Change from Baseline in PANSS Total Score Based on MMRM Analysis.

Study	Primary Endpoint	Lurasidone 40 mg	Lurasidone 80 mg	Lurasidone 120 mg	Olanzapine 15 mg
D1050006	PANSS (MMRM)	-14.1 p=0.009 (unadjusted)	NA	-16.2 p=0.0027 (unadjusted)	NA
D1050196	PANSS (MMRM)	NA	-10.2 p=0.0064	NA	NA
D1050229	PANSS (MMRM)	-2.1 p=0.591 (Hommel-based)	-6.4 p=0.034 (Hommel-based)	-3.5 p=0.391 (Hommel-based)	NA
D1050231	PANSS (MMRM)	-9.7 p=0.02 (Hommel-based)	NA	-7.5 p=0.022 (Hommel-based)	-12.6 p<0.001

Source: Clinical study reports D1050229 and D1050231, and Reviewer's results

5.2 CONCLUSIONS AND RECOMMENDATIONS

The sponsor submitted results of four studies (D1050006, D1050196, D1050229, D1050231) in support of efficacy of lurasidone (fixed doses of 40 mg/day, 80 mg/day, and 120 mg/day) versus placebo for the treatment of schizophrenia (b) (4)

In the primary analysis of BPRS Total score (studies D1050006, D1050196) and PANSS Total score (studies D1050229, D1050231), adult patients with a primary diagnosis of schizophrenia on lurasidone (fixed doses of 40 mg, 80 mg, and 120 mg) were observed to show statistically significant improvement over patients in the placebo treatment group. However, the 120 mg dose did not seem to add additional benefit over the other two doses. The results from study D1050006 with 70% dropout rate and the results from study D1050196 with 50% dropout rate should be interpreted with extra caution.

CGI-S was the pre-specified key secondary endpoint in studies D1050229 and D1050231. Study D1050229 demonstrated a statistical significance with respect to this endpoint for dose 80 mg, and study D1050231 for both doses 40 mg and 120 mg. CGI-S was declared as a secondary endpoint, but not pre-specified as a key secondary endpoint, in studies D1050006 and D1050196. Nevertheless, in study D1050006 the p-values from both doses (40 mg and 120 mg) were very small so that any multiple testing procedure would lead to a statistical significance for both doses. Study D1050196 investigated 80 mg only, and the p-value from testing this endpoint was also very small. Overall, lurasidone in doses 40 mg, 80 mg, and 120 mg was statistically superior to placebo in change from baseline in CGI-S score at week 6. However, whether the magnitude of improvement was of clinical relevancy is deferred to the clinical review team.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-200603	ORIG-1	SEPRACOR INC	Lurasidone HCl

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

/s/

GEORGE KORDZAKHIA
09/07/2010

PEILING YANG
09/07/2010

KOOROS MAHJOOB
09/07/2010
I discused the review with the primary reviewer and the Team leader. All views are incorporated in this version and I concur with it.

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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	200,603 / Serial 0000
Drug Name:	Lurasidone HCl Tablet
Indication:	Schizophrenia
Applicant:	Dainippon Sumitomo Pharma America, Inc. Fort Lee, New Jersey
Testing Facility:	(b) (4)
Date:	Submitted 30 December 2009
	To reviewer 6 February 2010
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Psychiatry Products
Toxicologist:	Sonia Tabacova, Ph.D.
Project Manager:	Ann J. Sohn
Keywords:	Carcinogenicity, Cox regresson, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY

The Sponsor's reports indicate that the objectives of these studies were to investigate the carcinogenic potential of the test article, Lurasidone HCl, code name SM-13496, in both rats and mice, as well as to assess any associated toxicokinetics. This review does not address the latter objective.

1.1. Conclusions and Recommendations

This submission summarizes the results of a two year rat study and a two year mouse study to assess the carcinogenic potential of Lurasidone HCl by daily oral gavage. Other gross aspects of the designs of the rat and mouse study are summarized in the following tables, for each gender in each species:

Table 1. Design of Rat Study (dose volume: 5 mL/kg)

Treatment Group	# animals	Lurasidone HCl Dosage (mg/kg/day)	Concentration (mg/mL) [1]
1. Vehicle	65	0	0
2. Vehicle	65	0	0
3. Low	65	3	0.6
4. Medium	65	12	2.4
5. High	65	50/36[1]	10/7.2[1]

[1]From days 1 through 403 in males and 1 to 402 in females, animals were dosed at a nominal 50 mg/kg/day (concentration 10 mg/mL).. On day 404 in males and 403 in females this was reduced to 36 mg/kg/day (concentration 7.2 mg/mL).

Note that both studies had nominally identical control groups. Please see Section 1.3.1.1 below for a comment on the statistical analysis.

Table 2. Design of Mouse Study (dose volume: 10 mL/kg)

Treatment Group	# animals	Dosage (mg/kg/day)	Concentration (mg/mL)
1. Vehicle	60	0	0
2. Vehicle	60	0	0
3. Low	60	30	3
4. Medium	60	100	10
5. Medium-High[1]	60	300	30
6. High	60	1200(M)[2] 650(F)	120(M) 65(F)

[1]The Sponsor labels this group as "Mid-High " and the preceding group as "Mid-Low".

[2] Up to day 410 males were dosed at 1200 mg/kg/day. Beginning on day 410 they were dosed at 650 mg/kg/day (as with female mice).

More detailed descriptions of the studies are provided in Sections 3.2.1 and 3.2.2.

The following table summarizes the survival results using the Sponsor supplied mortality data:

Table 3. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1&2, 3-5	0.0243	0.0204	0.4451	0.7525
Trend over Groups 1&2, 3-5	0.0131	0.0060	0.2749	0.5144
Homogeneity over Groups 1&2 vs 5	0.0332	0.0282	0.1639	0.4052
Mice Homogeneity over Groups 1&2, 3-6	0.3264	0.2464	0.0033	0.0045
Trend over Groups 1&2, 3-6	0.0429	0.0429	0.0001	0.0005
Homogeneity over Groups 1&2 vs 6	0.0404	0.0260	0.0006	0.0006

Figures A.1.1 through A.1.4, in Appendix 1, provide survival curves for each of the four species by gender combinations. These plots distinguish between the two control groups, but for the test results summarized above, the control groups are pooled. The trend test assesses the evidence for a dose related trend, either increasing or decreasing. From Figure A.1.1 for male rats there does seem to be some evidence of a dose related differences, but this is apparently a situation where, after an initial period, the high dose group actually seems to have the highest survival, i.e., lowest mortality (for all six tests, $0.0006 \leq p \leq 0.0332$). The low dose group generally has the lowest survival, but generally intertwined with those of the remaining groups. In female rats the corresponding Figure A.1.2 indicates that the survival curves of the dose groups are all quite intertwined, consistent with the hypothesis of no heterogeneity, trend, or differences between the high dose and pooled controls (all six $p \geq 0.1639$).

Unlike the situation with rats, in mice there is some evidence of a general decrement to survival associated with Lurasidone HCl. From figures A.1.3 and A.1.4 in Appendix 1, in both mouse genders there is a general decrease in survival over doses, with the highest mortality generally in the high dose group. Results are somewhat equivocal in male mice since the overall test for lack of homogeneity is not statistically significant (logrank $p = 0.3264$, Wilcoxon $p = 0.2464$), but with some evidence of both a trend (logrank and Wilcoxon $p = 0.0429$) and a difference between the high dose group and pooled controls (logrank $p = 0.0404$, Wilcoxon $p = 0.0260$). For female mice results are much stronger. In Figure A.1.4, the high dose group in female mice also has the highest mortality, closely followed by the medium high dose group. The remaining dose groups are generally intertwined. This explains the statistically significant results when testing for overall lack of homogeneity, trend, and no pairwise difference between the high dose group and pooled controls (all six $p \leq 0.0045$).

The significance levels of the tests of tumorigenicity in the FDA analysis are based on poly-k tests applied to the data sets provided by the Sponsor. The poly-k test modifies the

original Cochran-Armitage test of dose related trend in the occurrence of an event to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). One problem with any such tumorgenicity analyses is that for each tumor-organ-gender-study combination there is one test of significance for each comparison of an actual treatment group to controls plus a test of overall trend. This implies a large number of tests, necessitating a multiplicity adjustment. For two species, two gender per species studies the so-called Haseman-Lin-Rahman rules adjust for the multiplicity of tests of tumorigenicity by modifying the interpretation of the usual significance level (i.e. "p-value"). These specify that for tests of trend at a roughly overall 0.10 (10%) false positive error rate, one might claim statistical significance if the observed significance level is 0.005 for rare tumors (with a historical control incidence less than 1%) and 0.025 (incidence at or greater than 1%) for common tumors. Tests comparing the high dose group to controls would be considered statistically significant if the observed significance level is 0.05 for rare tumors and 0.01 for common tumors. This adjustment for multiplicity is discussed in Section 1.3.1.5 below. Tables 4 and 5 below, display the tumor incidence in both rats and mice, respectively, as well as the results of tests of no differences between treatments for those neoplasms that had at least one test that achieved a nominal 0.05 level of significance. Note again that while tumor incidence is tabulated separately for the two vehicle groups, the actual carcinogenicity tests also utilize pooled vehicle groups.

Table 4. Potentially Statistically Significant Neoplasms in Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low vs Veh
N	65	65	65	65	65				
Male Rats									
HEMATO NEOPLASIA									
M-LYMPHOMA	0	1	4	1	2	0.4121	0.3047	0.5432	0.0396
MAMMARY, MALE									
Adenoma/Carc./Fibro.	0	1	1	4	1	0.5059	0.6019	0.0372	0.5284
SKIN									
Fibroma/Fibrosarcoma	2	0	2	7	1	0.6429	0.2982	0.0064	0.3789
M-FIBROSARCOMA	1	0	2	5	0	0.8026	0.3630	0.0156	0.2364
Female Rats									
MAMMARY, FEMALE									
Adenoma/Carc./Fibro./mixed	38	38	41	50	42	0.1771	0.1966	0.0121	0.3523
B-ADENOMA	11	5	10	20	12	0.2538	0.2477	0.0044	0.4292
M-CARCINOMA	19	14	21	30	32	0.0008	0.0009	0.0027	0.2483
THYROID									
Adenoma/Carc. C cell	4	4	2	4	8	0.0398	0.1423	0.3807	0.7363

In Table 4, above, in female rats, following the adjustment for multiplicity to get an overall rough 10% error rate and using the incidence in the no treatment group to decide if a tumor is rare or not, we would conclude, that the test of trend in malignant mammary carcinoma was statistically significant ($p = 0.0008 < 0.005$), as was the test comparing the high dose to the pooled vehicle ($p = 0.0009 < 0.01$). No other tests of trend or comparisons between the high dose and controls achieved the multiplicity adjusted significance levels. Applying the rule for pairwise comparisons to other groups than the high dose group can be expected to inflate the type I error rate above the rough 10% level established in Lin and Rahman (2006).

But if such a rule is used the comparison beween the low dose and vehicle in hemato neoplasia lymphoma in male rats would be statistically significant, though barely ($p = 0.0372 < 0.05$), as would be the comparison between the vehicle and medium dose groups for pooled adenoma and other neoplasms in the mammarys of male rats ($p = 0.0396 < 0.05$). In male rats the tests of differences between the medium dose group and pooled controls in skin fibroma/fibrosarcoma would also be statistically significant ($p = 0.0064 < 0.01$), as would be the test of malignant fibrosarcoma ($p = 0.0156 < 0.05$). Again accepting the probable inflation of overall Type I error, we would also note that the pairwise comparisons between the medium dose group and pooled controls in mammary adenoma and mammary carcinoma would also be classified as being statistically significant ($p = 0.0044 < 0.01$, $p = 0.0027 < 0.01$, respectively). No other comparisons would meet these looser criteria.

Table 5. Potentially Statistically Significant Neoplasms in Mice (All Female)

	Veh 1	Veh 2	Low	Mid-Med	Hi	High Hi	Trend	High vs Veh	Med-Hi vs Veh	Medium vs Veh	Low vs Veh
N	60	60	60	60	60	60					
ADRENAL, MEDULLA											
B-PHEOCHROMOCYTOMA	0	0	0	0	3	0	0.3776	.	0.0261	.	.
HARDERIAN GLAND											
Adenoma/Carcinoma	5	3	7	3	5	8	0.0354	0.0489	0.3406	0.5383	0.1383
MAMMARY, FEMALE											
Adenoma/Carc./-sarcoma/-canth.	2	1	13	19	26	20	0.0000	0.0000	0.0000	0.0000	0.0000
M-ADENOACANTHOMA	1	0	7	6	7	5	0.1028	0.0080	0.0011	0.0061	0.0014
M-CARCINOMA	2	1	7	12	18	13	0.0011	0.0000	0.0000	0.0002	0.0113
M-CARCINOSARCOMA	0	0	0	1	2	2	0.0259	0.0800	0.0898	0.3333	.
OVARY											
Cystad./Gran./Thecal/Tubul.	1	0	2	4	1	2	0.2543	0.1961	0.5073	0.0446	0.2307
PANCREAS											
B-ISLET CELL ADENOMA	0	0	0	1	3	1	0.1546	0.2857	0.0261	0.3333	.
PITUITARY											
B-ADENOMA, PARS DISTALIS	3	4	11	17	27	29	0.0000	0.0000	0.0000	0.0001	0.0068

In male mice, no tests of trend or tests of pairwise differences from the control achieved even the nominal 0.05 level of significance, let alone the levels adjusted for multiplicity. However, in female mice there is some strong evidence of a drug effect on mammary and pituitary tumors. In particular, the tests of overall trend and pairwise comparison between the high dose group and control in mammary malignant carcinoma in females was statistically significant ($p = 0.0011 < 0.005$ and $p < 0.00005 < 0.01$, respectively), as were the tests of pooled tumors ($p < 0.00005 < 0.005$ and $p < 0.00005 < 0.01$). Similarly the tests of overall trend and pairwise comparison between the high dose group and control in pituitary pars distalis adenoma in females was highly statistically significant (both $p < 0.00005 < 0.005$ and 0.01, respectively). Note that while the relative constancy of mammary adenoacanthoma across the actual Lurasidone treatment groups implies there is no strong evidence of trend, the comparison the high dose to control is statistically significant ($p = 0.008 < 0.05$). Again, incorporating the other pairwise comparisons can be expected to raise rhe nominal type I error rate to something above the rough 10% level. But if one accepts that potential inflation, the pairwise comparisons of the medium-high and medium dose groups in mammary malignant carcinoma were statistically significant ($p < 0.00005 < 0.01$ and $p = 0.0002 < 0.01$, respectively), while comparison in the low dose group was close to adjusted statistical

significance ($p = 0.0113 \approx 0.01$). However, in pooled mammary tumors all these comparisons would also be statistically significant (all three $p < 0.00005 < 0.01$). These pairwise comparisons in mammary adenocanthoma would also be labeled as statistically significant (all $p = 0.0011, 0.0061, 0.0014 < 0.05$). Similarly the pairwise comparisons of the medium-high, medium, and low dose groups to the pooled controls in pituitary pars distalis adenoma would also be classified as statistically significant (all three $p < 0.00005$, $p = 0.0001$, $p = 0.0068 < 0.01$). The comparison between the medium dose group and pooled vehicle in pooled cancers of the ovary would classified as statistically significant ($p = 0.0446 < 0.05$), though only barely. Finally, the comparisons of the medium-high dose group to the the pooled vehicle in adrenal pheochromocytoma and islet cell adenoma of the pancrease were statistically significant ($p = 0.0261 < 0.05$). Following the Haseman-Lin-Rahman rules above, no other tests achieved statistical significance, though several were close.

1.2. Brief Overview of the Studies

This submission had a standard rat study:

Protocol 6645-139: 104-Week Oral Gavage Carcinogenicity and Toxicokinetic Study with SM-13496 in Rats,

and a standard mouse study :

Protocol 6645-138: 104-Week Oral Gavage Carcinogenicity and Toxicokinetic Study with SM-13496 in Mice

to assess the carcinogenic potential of Lurasidone HCl in rodents.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

In this section, several issues, typical of statistical analyses of these studies, are considered. These issues include the usage of dual control groups, and details on the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Dual Controls:

The Sponsor provides two supposedly identical vehicle control groups in each study. All tables and plots in this report distinguish between the two control groups, groups 1 and 2. The Sponsor states that: “For each specific tumor type, Group 1 will be compared with Group 2 with significance declared as follows (common tumors; $p \leq 0.01$; rare tumors; $p \leq 0.05$). If there are no differences detected between Groups 1 and 2, then data from Group 1 and 2 will be combined as one control group, and tests for effects of the test article will be assessed with trend comparisons (common tumors; $p \leq 0.005$; rare tumors; $p \leq 0.025$), or group comparisons

($p \leq 0.01$; rare tumors; $p \leq 0.05$), as appropriate to the data set and analysis being done. If there are differences detected between Groups 1 and 2, then the tests for the effects of the test article will be assessed with each control group separately . . ." (pages 335-336 of rat report, pages 18-19 of protocol)

The first issue with this procedure is that results of tests on treatment groups are conditional on the outcomes of the tests between the controls, whereas the significance values are computed assuming the tests are not conditional. Thus the distributional assumptions of the usual unconditional tests are not met. Also, of more importance is that unless there are systemic problems with the conduct of the study, any observed differences should be due to random fluctuations between the treatment groups. That is, pre-study randomization to two identical controls should be equivalent to post-study randomization into two control groups. In the latter circumstances it would seem that few analysts would place any weight on observed differences between the control groups (since a simple rerandomization would almost surely eliminate any differences). But then logically no weight should be placed on any observed differences between vehicle controls in the current studies, and on differing results when control groups are tested against other treatment groups. Finally, note that this procedure increases the number of statistical tests, and thus increases the probability of a false conclusion of treatment differences. Hence, this reviewer would argue against the separate analyses as provided by the Sponsor. For these reasons, all tests in the FDA analysis, both tests of differences in survival and tests of differences tumorigenicity use a single pooled control group and ignore possible differences in controls.

1.3.1.2. Survival Analysis:

The survival analyses presented here are based on both the log rank test and the Wilcoxon test comparing survival curves. The log rank tests tend to puts higher weight on later events, while the Wilcoxon test tends to weight events more equally, and thus is more sensitive to earlier differences in survival. The log rank test is most powerful when the survival curves track each other, and thus the proportional hazard assumption seems to be true. Both tests were used to test both homogeneity of survival among the treatment groups and the effect of dose on trend in survival. Appendix 1 reviews the specific animal survival analyses in more detail. The results of the similar Sponsor's analyses are summarized in Sections 3.2.1.1 and 3.2.2.1.

1.3.1.3. Multiplicity of Tests on Survival:

Using the logrank and Wilcoxon tests, there are six tests of survival in each species by gender combination. If we were to assume the tests are independent across comparisons, which clearly they are not, and assume that there is absolutely no difference in survival, the probability of at least one statistically significant result in each species, at the usual 0.05 level, is about 0.4596. Such is the possible price paid for the multiplicity of hypothesis tests.

1.3.1.4. Tests on Neoplasms:

The Sponsor's reports states that incidental tumors (i.e., tumors not assigned to be the cause of death of the animals by the study pathologist) were analyzed by linear logistic

regression of tumor prevalence. Fatal and mortality independent (palpable) tumors were analyzed by a binary regression method using the death time or time of detection as a surrogate for the tumor onset time. Results were pooled as in a standard Peto type analysis. In rats: "Since the dose level for the high-dose group was changed during the study, ordinal Dose Levels 0, 1, 2, 3, and 4 were used in all analyses for Groups 1, 2, 3, 4, and 5, respectively. Continuity correction was done for all asymptotic tests." (page 450 of rat report)

Appendix 2 presents the results from the FDA poly-k analysis on tumor incidence in rats and mice. The poly-k test is a modification of the original Cochran-Armitage test of trend in response to dose, adjusted for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). For rats the Sponsor provided the results of so-called Peto tests of carcinogenicity, here applied to pairwise differences with the vehicle control. These tests require accurate specification of the cause of death, which is often difficult. It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that the poly-k modification of the Cochran-Armitage tests of trend has been recommended over the corresponding Peto tests.

Also note that in rats, the Sponsor's dose weighting apparently assumes a one unit difference between the nominally identical control groups, plus equal increments in dose effect over increasing treatments. These could be expected to have a downweighting effect on results. The FDA analysis follows the intent-to-treat principle and uses treatment weight as initially randomized.

1.3.1.5. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms necessitates a number of statistical tests, which in turn necessitates an adjustment in experiment-wise Type I error (i.e., the probability of rejecting a true null hypothesis). Based on his extensive experience with such carcinogenicity analyses in standard laboratory rodents, for pairwise tests between the high dose group and controls in two species, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Similarly, Lin and Rahman (1998) showed that tests of trend should be tested at a 0.025 level for rare tumors and 0.005 for common tumors. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

1.3.1.6. Validity of the Designs:

When determining the validity of designs there are two key points:

1) adequate drug exposure,
2) tumor challenge to the tested animals.

1) is related to whether or not sufficient animals survived long enough to be at risk of forming late-developing tumors and 2) is related to the Maximum Tolerated Dose (MTD), designed to achieve the greatest likelihood of tumorigenicity.

Lin and Ali (2006), quoting work by Haseman, have suggested that in standard laboratory rodent species, a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. Note that as a percentage of animals that survived to week 91, this criterion is met in rats in all dose groups in both genders, and in male mice. Only for the medium-high dose group and the high dose group in female mice is this criterion exceeded. (Please see table 9 on page 15, table 10 on page 16, table 14 on page 20, and table 15 page 21).

The mean weight values in the following tables were taken from the Sponsor's rat and mice reports (Table 12, Rat study pages 108-119, Table 7, Mice study pages 83-93). The change from baseline is the simple difference between means and is not mortality adjusted.

Table 6. Mean Weights for Mice and Rats

Rats Dose Group	Males				Females			
	Week		Change from baseline	% change relative to pooled vehicle	Week		Change from baseline	% change relative to pooled vehicle
	1	105			1	100		
Vehicle 1	206	806			173	530		
Vehicle 2	207	762			171	600		
Pooled Veh	206.5	781.8	575.3		172	547.8	375.8	
Low	206	735	529	92.0%	173	544	371	98.7%
Medium	206	680	474	82.4%	172	418	246	65.5%
High	206	608	402	69.9%	171	400	229	60.9%

Mice Dose Group	Males				Females			
	Week		Change from baseline	% change relative to pooled vehicle	Week		Change from baseline	% change relative to pooled vehicle
	1	105			1	98		
Vehicle 1	29.6	45.3			24.3	40.3		
Vehicle 2	29.6	46.2			24.1	39.9		
Pooled Veh	29.6	45.8	16.2		24.2	40.1	15.9	
Low	29.7	43.0	13.3	82.1%	24.0	41.5	17.5	110.1%
Medium	29.3	44.2	14.9	92.0%	24.1	39.0	14.9	93.7%
Med-High	29.8	42.6	12.8	79.0%	24.0	39.6	15.6	98.1%
High	29.3	41.3	12.0	74.1%	24.0	38.5	14. 5	91.2%

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag *et al* (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' " From Table 6, above, in

both rat genders and in male mice there is a clear decrement in the high dose Lurasidone groups compared to vehicle, with an apparent trend of decreasing weight over increasing dose. This may explain at least part of the observed lower mortality in the high dose group in male rats.

In male and female rats, the Sponsor summarizes overall food consumption as follows: "[T]he values for most intervals were similar. In males given 3 or 12 mg/kg/day, there were no consistent effects on food consumption. In males given, 50/36 mg/kg/day, food consumption was reduced in comparison to each of the control groups for most of the measured intervals." (page 22 of rat report)

The Sponsor concludes that: "Overall, in females given 3 mg/kg/day, food consumption was increased in comparison to controls. Conversely, in females given 12 mg/kg/day, and in males and females given 50/36-mg/kg/day, food consumption was reduced in comparison to controls. At the dose level of 50/36 mg/kg/day, the lower food consumption values were consistent with the lower body weights and body weight gains for this group." (page 22 of report)

In mice the Sponsor states that "there was no evidence of consistent dose or test article-related changes in food consumption. Total food consumption (from Day 1 to terminal sacrifice) in the females was statistically significantly decreased in Groups 4, 5, and 6 compared to the vehicle control. The decreases in female total food consumption were comparable in all dose levels (2.0 to 9.0% lower than the vehicle control) and did not occur in a dose-related manner, but may have been associated with treatment. There were no statistically significant differences in total food consumption in the test article-treated males when compared to the vehicle control group (Group 1)." (page 30 of report) Whether these observations have any effect on the assessment of the nominal MTD is a decision requiring the expertise of the toxicologist.

Again from 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. If dosing is close to the MTD one would expect slightly higher mortality due to toxicity, but not so much that it largely reduces the number of animals exposed to the drug. In male rats there is evidence of lower mortality in the high dose group compared to the other groups. In female rats all actual dose groups seem to have similar mortality, somewhat higher than the vehicle. In both mice genders, but particularly in females,mortality in the high dose group does seem to differ from the control groups. A related way to assess whether or not the MTD was achieved is to measure mortality not associated with any identified tumor. Table 7, below, indicates that the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors (i.e., the "Event"):

Table 7. Natural Death with No Identified Tumor

Rats		1.Vehicle 0 mg/kg	2. Vehicle 0 mg/kg	3. Low 3 mg/kg	4.Medium 12 mg/kg	5. High 50/36 mg/kg
Male	Event	9	8	9	10	7
	No event	56	57	56	55	58
Female	Event	1	2	1	0	1
	No event	64	63	64	65	64

Mice		1.Vehicle 0 mg/kg	2. Vehicle 0 mg/kg	3. Low 30 mg/kg	4.Medium 100 mg/kg	5. Med-Hi 300 mg/kg	6. High 1200 mg/kg [1]
Male	Event	16	16	23	24	28	22
	No event	44	44	37	36	32	38
Female	Event	16	10	11	11	8	12
	No event	44	50	49	49	52	48

[1]In female mice this dosage was 650 mg/kg/day

Clearly in rats there is no evidence of treatment related differences in natural death prior to tumor development (i.e., the "Event" above). For a frequentist test of hypotheses of no differences we can specify the usual survival tests where animals that die with a tumor or are sacrificed are considered as censored. The remaining animals are those that die a natural death prior to developing a tumor. One indication that the MTD is achieved would be dose related excess toxicity, resulting in a dose related increase in these deaths, particularly in the high dose group. In female rats the event incidence is probably too low to justify the asymptotics used to compute significance levels. Comparing the high dose to the pooled vehicle group in male rats only validates the rather obvious observation that there are no dose related differences (Male rats: log rank p = 0.5665, Wilcoxon p = 0.56836). In mice results are rather more equivocal (Male mice: log rank p = 0.0817, Wilcoxon p = 0.0565, Female mice: log rank p = 0.6481, Wilcoxon p = 0.3336). Although this is a decision for the toxicologist, this may be evidence that the MTD was not exceeded in rats and female mice, but, depending upon the significance level used, may have been slightly exceeded in male mice.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

This submission summarizes the results of a two year rat study and a two year mouse study to assess the carcinogenic potential of by daily oral gavage.

2.2. Data Sources

The Sponsor provided eight SAS data sets, four each for rats and mice, with the obvious contents.:

Rats: tumor.sas7bdat food.sas7bdat Mice: tumor.sas7bdat food.sas7bdat
mortal.sas7bdat weights.sas7bdat mortal.sas7bdat weights.sas7bdat

Only the tumor and mortality data were used in this report. Other cited values were taken from the Sponsor's report.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1. Protocol 6645-139: 104-Week Oral Gavage Carcinogenicity and Toxicokinetic Study with SM-13496 in Rats,

STUDY DURATION: 104 Weeks (99 weeks of treatment)
INLIFE START DATE: Males: 25 November 2003, Females: 26 November 2003
INLIFE END DATE: 30 November 2005
RAT STRAIN: (b) (4) Crl:CD®(SD)IGS BR Rats
ROUTE: Oral Gavage

The basic design of the rat study has five dose groups, summarized in Table 8 below:

Table 8. Design of Rat Study (dose volume: 5 mL/kg)

Group	Main Study # animals /gender	Toxicokin -etic # / gender	Dosage (mg/kg/day)	Concentration (mg/mL) [1]
1. Vehicle	65	15	0	0
2. Vehicle	65	-	0	0
3. Low	65	15	3	0.6
4. Medium	65	15	12	2.4
5. High	65	15	50/36[1]	10/7.2[1]

[1]From days 1 through 403 in males and 1 to 402 in females, animals were dosed at a nominal 50 mg/kg/day (concentration 10 mg/mL).. On day 404 in males and 403 in females the dosage of main study animals was reduced to 36 mg/kg/day (concentration 7.2 mg/mL).

The Sponsor states that animals were randomly allocated to treatment, apparently stratified by weight. Treatment was administered by oral gavage daily for at least 104 weeks at a dose volume of 5 mL/kg and was continued through the day prior to scheduled sacrifice. The Sponsor indicates that this an appropriate route of administration since humans are intended to use oral dosing of Lurasidone HCl. The two vehicle groups were dosed solely with the vehicle, i.e. 0.5% (w/v) medium viscosity (1500 cps) methylcellulose in reverse osmosis (RO) water. Animals were housed individually, with food and water available *ad libitum*.

The Sponsor states that: "At initiation of treatment, the animals were approximately 6 weeks old, and their body weights ranged from 167 to 241 g for the males and 144 to 199 g for the females. Following randomization, each study animal was assigned a unique number by means of an implantable microchip identification device and/or cage card. Animals not used on study were sacrificed and discarded." Further, "The dosages given for this study were based on the most recently recorded body weight." (pages 14-15 of report)

The Sponsor justifies dose levels as follows: " Dose selection for this study was based upon toxicologic findings during a 6-month oral gavage toxicity study with SM-13496 in rats (Sumitomo Chemical Co., Ltd., Study No. 3259) using dose levels of 0, 0.03, 1, 10, and 100 mg/kg/day. No test article-related mortalities were observed. Decreased spontaneous activity and ptosis were noted at 10 mg/kg/day and above. Significantly reduced mean body weight gain was observed for males and females (-23.5% for each sex) given 100 mg/kg/day." (page 15 of report)

3.2.1.1. Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigenicity in rats.

Survival analysis:

The Sponsor states that: "a significant negative trend was observed in mortality in the males (by Cox-Tarone and Gehan-Breslow tests versus Control 1 and by Gehan-Breslow test versus Control 2). This negative trend was strictly due to decreased mortality in the high-dose group. The two control groups indicated similar mortalities, and neither of the other two treated groups was dissimilar to the two controls."

The Sponsor continues: "in the females, no significant change in mortality was noted in any of the treated groups compared with either of the two control groups, although in general, all treated groups showed lower mortality rates compared with either of the two controls. The decreased mortality rates in the treated groups were not severe enough to show any statistical significance in females." (page 451 of report)

Tumorigenicity analysis:

The Sponsor summarizes carcinogenicity results as follows: "In males . . . the 12-mg of SM-13496/kg of body weight/day (mg/kg/day) group was significantly increased over Control

2 for skin/muscle, skeletal fibrosarcoma, and fibroma/ fibrosarcoma combined. No other significant effects, either in terms of positive trend or significant increase over the controls, were noted."

"In . . . females . . . significant negative trends for pituitary adenoma and adenoma/ carcinoma combined versus both controls were observed, with high-dose significant decrease over Control 2 ($p = 0.0026$ for adenoma and $p = 0.0081$ for adenoma/carcinoma combined). Statistically significant decreases in the 12-mg/kg/day group of pituitary adenoma and adenoma/carcinoma combined versus both controls were also noted. The 12-mg/kg/day group showed significant increase over both controls ($p = 0.0492$ and $p = 0.0021$, respectively) in mammary adenoma. In mammary carcinoma, significant positive trends versus Control 2 were observed, with high-dose significant increase over Control 2 ($p = 0.0025$). The 12-mg/kg/day group also showed significant increase over both controls ($p = 0.0277$ and $p = 0.0026$, respectively) for mammary carcinoma. When the adenoma, carcinoma, and fibroadenoma incidences were combined, the significant trend disappeared, but the 12-mg/kg/day group significant increase against the two controls remained ($p = 0.0286$ and $p = 0.0367$, respectively). No other statistically significant increase or decrease in the female neoplastic lesions was noted." (pages 451-452 of report).

3.2.1.2. FDA Reviewer's Results

This section will present the Agency findings on survival and tumorigenicity in male and female rats.

Survival analysis:

The following tables (Table 9 for male rats, Table 10 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. The Kaplan-Meier survival plots of in Appendix 1 provide a more detailed picture of mortality losses.

Table 9. Summary of Male Rats Survival (dosed at mg/kg/day)

Period (Weeks)	Vehicle 0	Vehicle 0	Low 3	Medium 12	High 50/36
1-52	4/65[1] 93.8%[2]	4/65 93.8%	4/65 93.8%	5/65 92.3%	2/65 96.9%
53-78	11/61 76.9%	15/61 70.8%	18/61 66.1%	10/60 76.9%	9/63 83.1%
79-91	16/60 52.3%	11/46 53.8%	15/43 43.1%	19/50 47.6%	6/44 73.8%
92-104	16/44 27.7%	13/35 33.9%	10/28 27.7%	17/31 21.5%	19/24 44.6%
Terminal 105	18	22	18	14	29

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.

In these tables the terminal period only includes those animals that were sacrificed. Animals that died of other causes during the terminal period are included in the preceding time period.

Table 10. Summary of Female Rats Survival (dosed at mg/kg/day)

Period (Weeks)	Vehicle 0	Vehicle 0	Low 3	Medium 12	High 50/36
1-52	1/65[1] 98.5%[2]	3/65 95.4%	2/65 96.9%	3/65 95.4%	3/65 95.4%
53-78	19/64 69.2%	18/62 67.7%	17/63 70.8%	17/62 69.2%	17/62 69.2%
79-91	13/45 49.2%	18/44 40.0%	16/46 46.1%	13/45 49.2%	12/45 50.8%
92-104	17/32 23.1%	14/26 18..5%	10/30 30.8%	16/32 24.6%	12/33 32.3%
Terminal 105	15	12	20	16	21

[1] number of deaths / number at risk
[2] per cent survival to end of period.

Table 11 below provides the significance levels of the tests of homogeneity and trend over dose groups as proposed in Section 1.3.1.1, above.

Table 11. Statistical Significances of Tests of Homogeneity and Trend in Survival in Rats

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1&2, 3-5	0.0243	0.0204	0.4451	0.7525
Trend over Groups 1&2, 3-5	0.0131	0.0060	0.2749	0.5144
Homogeneity over Groups 1&2 vs 5	0.0332	0.0282	0.1639	0.4052

In Figure A.1.1 in Appendix 1, for male rats there does seem to be evidence of dose related differences, but after a starting period, the high dose group has the lowest mortality (i.e., greatest survival), clearly separated from the remaining groups. Meanwhile, the low dose group generally has the highest mortality, although the survival curves of the remaining dose groups are generally closely intertwined with the survival curve of this group. This is consistent with the results of the various statistically significant tests in male rats (Overall homogeneity: LR p = 0.0243, Wilcoxon p = 0.0204, Trend: LR p = 0.0131, Wilcoxon p = 0.0006, High vs. control: LR p = 0.0332, Wilcoxon p = 0.0282). From Figure A.1.2 below, in female rats it seems that the the dose groups are all quite intertwined, consistent with the hypothesis of no heterogeneity, trend, or differences between the high dose and pooled controls (all six p ≥ 0.1639).

Tumorigenicity analysis:

As discussed in Section 1.3.1.5, for common tumors, the Haseman-Lin-Rahman rules are that for a roughly 0.10 (10%) overall false positive error rate, overall trend should be tested at a 0.025 (2.5%) level in rare tumors and at 0.005 (0.5%) in common tumors. Pairwise tests between the high dose group and control should be tested at 0.05 (5%) level in rare tumors and at a 0.01 (1%) in common tumors. Table 12 below lists those organ by tumor combinations that have at least one test of trend and pairwise comparisons with a nominal significance level of 0.05.

Table 12. Potentially Statistically Significant Neoplasms in Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low ve Veh
N	65	65	65	65	65				
Male Rats									
HEMATO NEOPLASIA									
M-LYMPHOMA	0	1	4	1	2	0.4121	0.3047	0.5432	0.0396
MAMMARY, MALE									
Adenoma/Carc./Fibro.	0	1	1	4	1	0.5059	0.6019	0.0372	0.5284
SKIN									
Fibroma/Fibrosarcoma	2	0	2	7	1	0.6429	0.2982	0.0064	0.3789
M-FIBROSARCOMA	1	0	2	5	0	0.8026	0.3630	0.0156	0.2364
Female Rats									
MAMMARY, FEMALE									
Adenoma/Carc./Fibro./mixed	38	38	41	50	42	0.1771	0.1966	0.0121	0.3523
B-ADENOMA	11	5	10	20	12	0.2538	0.2477	0.0044	0.4292
M-CARCINOMA	19	14	21	30	32	0.0008	0.0009	0.0027	0.2483
THYROID									
Adenoma/Carc. C cell	4	4	2	4	8	0.0398	0.1423	0.3807	0.7363

In Table 12, above, in female rats the test of trend in malignant mammary carcinoma was statistically significant ($p = 0.0008 < 0.005$), as was the test comparing the high dose to the pooled vehicle ($p = 0.0009 < 0.01$). No other tests of trend or comparisons between the high dose and controls achieved the multiplicity adjusted significance levels in male or female rats. Again, accepting the inflation of type I error rate above the rough 10% level, the comparison between the low dose and vehicle in hemato neoplasia lymphoma in male rats would also be statistically significant, though barely ($p = 0.0396 < 0.05$), as would be the comparison between the vehicle and medium dose groups in terms of pooled adenoma and other neoplasms in the mammarys of male rats ($p = 0.0372 < 0.05$). In male rats the test of difference between the medium dose group and pooled controls in skin fibroma/fibrosarcoma would also be statistically significant ($p = 0.0064 < 0.01$). We would also note that the pairwise comparisons between the medium dose group and pooled controls in both mammary adenoma and mammary carcinoma would also be classified as being statistically significant ($p = 0.0044 < 0.01$ and $p = 0.0027 < 0.01$, respectively). No other comparisons would meet these looser criteria.

3.2.2. Protocol 6645-138: 104-Week Oral Gavage Carcinogenicity and Toxicokinetic Study with SM-13496 in Mice

STUDY DURATION: Males: 104 Weeks, Females: 98 Weeks
INLIFE START DATE: Males: 8 March 2004, Females: 11 March 2004
INLIFE END DATE: 10 March 2006
RAT STRAIN: (b) (4) Crl:CD®(SD)IGS BR Mice
ROUTE: Oral Gavage

The basic design of the mouse study has six dose groups, summarized in Table 8 below:

Table 13. Design of Mouse Study (Daily gavage: dose volume 10 mL/kg)

Group	Males Dosage (mg/kg)	Males Concen-tration (mg/mL)	Females Dosage (mg/kg)	Females Concen-tration (mg/mL)	# animals Main Study	# animals Toxico-kinetic	# animals Prolactin
1. Vehicle	0	0	0	0	60	NA	10
2. Vehicle	0	0	0	0	60	NA	NA
3. Low	30	3	30	3	60	39[2]	10
4. Medium	100	10	100	10	60	39[2]	10
5. Med-High[1]	300	30	300	30	60	39[2]	10
6. High	1200/650[2]	120/65[2]	650	65	60	39[2]	10

[1] The Sponsor labels this group as "Mid-High " and the preceding group as "Mid-Low".
[2] Up to day 410 males were dosed at 1200 mg/kg/day. Beginning on day 410 they were dosed at 650 mg/kg/day (as with high dose female mice).

The Sponsor states that animals were randomly allocated to treatment, apparently stratified by weight. Treatment was administered by oral gavage daily planned for at least 104

weeks at a dose volume of 10 mL/kg and was continued through the day prior to scheduled sacrifice. The two vehicle groups were dosed solely with the vehicle, i.e. 0.5% (w/v) medium viscosity (1500 cps) methylcellulose in reverse osmosis (RO) water. Animals were housed individually, with food and water available *ad libitum*.

The Sponsor justifies dose levels as follows: " Dose selection for this study was based in part upon toxicologic findings during (b) (4) 6645-135, '14-Day Oral Gavage Toxicity Study with SM-13496 in Mice' using dose levels of 0, 100, 300, and 1000 mg/kg/day and (b) (4) 6645-136, '13-Week Oral Gavage Preliminary Carcinogenicity and Toxicokinetic Study with SM-13496 in Mice' using dose levels of 0, 25, 125, 250, and 500 mg/kg/day." (page 17 of report) The Sponsor indicates that there were no test article-related mortalities in either study, but there were some small weight gain decrements in the high dose groups.

3.2.1.1. Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigenicity in rats.

Survival analysis:

The Sponsor states that: "no statistically significant trends or increases in the treated group mortality rates was noted in the males. The two controls in males were similar in mortality rates." (page 3912 of mouse report)

In females, "statistically significant positive trends and increases ub animals given 300 or 650 mg of SM-13496/kg of body weight/day (mg/kg/day) were noted in survival rates by both the Cox-Tarone and Gehan-Breslow tests. The two female controls were statistically similar in their mortality rates." (page 3912 of mouse report)

Tumorigenicity analysis:

The Sponsor summarizes carcinogenicity results as follows: "[I]n males, a significant decrease was noted in adrenal cortex subcapsular cell adenoma in animals given the 300 mg/kg/day versus Control Group 1 (p=0.0305). No such effect was found versus Control Group 2. Significant decreases for liver hepatocellular adenoma, multiple and carcinoma combined were observed for animals given 300 or 1200/650 mg/kg/day (p=0.0207 and p=0.0072) versus Control Group 1 and animals given 300 mg/kg/day (p=0.0182) versus Control Group 2. A significant negative trend (p=0.0012) was noted in lung bronchiolar/alveolar adenoma, multiple combined with carcinoma, multiple versus Control Group 2, with significant decreases in animals given 30 (p=0.0081), 100 (p=0.0195), 300 (p=0.0013), or 1200/650 (p=0.0024) mg/kg/day. Significant decreases were also noted in animals given 30 (p=0.0378) or 300 (p=0.0070) mg/kg/day versus both control groups (p=0.0369 and p=0.0227, respectively) for lung multiple bronchiolar/alveolar carcinoma. No other significant effects, either in terms of positive/negative trend or significant increase/decrease over controls, were noted."

"In females, . . . significant positive trends for pituitary adenoma, pars distalis versus both controls were observed (p=0.0000 in both cases), with significant increases noted in animals given 30 (p=0.0162 versus Control Group 1 and p=0.0373 versus Control Group 2), 100 (p=0.0195), 300 (p=0.0000 in both cases), or 650 (p=0.0000 in both cases) versus both control groups. Significant increases were observed in mammary malignant adenoacanthoma in animals given 30 (p=0.0236) or 300 (p=0.0155) mg/kg/day over Control Group 1 and in animals given 30 (p=0.0072), 100 (p=0.0178), 300 (p=0.0044), or 650 (p=0.0083) over Control Group 2. In mammary carcinoma, significant positive trends versus both controls (p = 0.0000 in both cases), with significant increases in animals given 100 (p=0.0073), 300 (p=0.0001), or 650 (p=0.0004) mg/kg/day over Control Group 1 and animals given 30 (p=0.0257), 100 (p=0.0025), 300 (p=0.0000), or 650 (p=0.0001) mg/kg/day over Control Group 2.. When the mammary malignant carcinoma and adenocanthoma were combined, significant positive trends both controls were observed (p = 0.0000 in both cases). Significant increases were noted in all treated groups (p=0.0016, p=0.0002, p=0.0000, and p=0.0000, respectively, versus Control Group 1 and p=0.0007, p=0.0001, p=0.0000, and p=0.0000, respectively, versus Control Group 2). No other significant increases or decreases in female neoplastic lesions were noted." (pages 3912-3913 of mice report).

Again, this reviewer considers some aspects of the Sponsor's analysis to be somewhat problematic (please see Sections 1.3.1.1 and 1.3.1.4 for discussion).

3.2.1.2. FDA Reviewer's Results

This section will present the Agency findings on survival and tumorigenicity in male and female rats.

Survival analysis:

The following tables (Table 14 for male mice, Table 15 for female mice) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. The Kaplan-Meier survival plots of in Appendix 1 provide a more detailed picture of mortality losses.

Table 14. Summary of Male Mice Survival (dosed at mg/kg/day)

Period (Weeks)	Vehicle 1 0	Vehicle 2 0	Low 30	Medium 100	Medium-High 300	High 1200
1-52	2/60[1] 96.7%[2]	5/60 91.7%	5/60 91.7%	6/60 90.0%	5/60 91.7%	12/60 80.0%
53-78	4/58 76.9%	11/55 73.3%	10/55 75.0%	9/54 75.0%	12/55 71.7%	8/48 66.7%
79-91	15/54 65.0%	8/44 60.0%	8/45 61.7%	9/45 60.0%	10/43 65.0%	12/40 46.7%
92-104	12/39 45.0%	7/36 48.3%	13/37 40.0%	14/36 36.7%	10/33 38.7%	8/28 33.3%
Terminal 105	27	29	24	22	23	20

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.

In these tables the terminal period only includes those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period.

Table 15. Summary of Female Mice Survival (dosed at mg/kg/day)

Period (Weeks)	Vehicle 1 0	Vehicle 2 0	Low 30	Medium 100	Medium-High 300	High 650
1-52	4/60[1] 93.3%[2]	2/60 96.7%	6/60 90.0%	3/60 90.0%	6/60 90.0%	6/60 90.0%
53-78	11/56 75.0%	12/58 76.7%	17/54 61.7%	11/57 76.7%	15/54 65.0%	20/54 56.7%
79-91	12/45 55.0%	16/46 50.0%	7/37 50.0%	12/46 56.7%	22/39 28.3%	15/34 31.7%
92-97	8/33 41.7%	10/30 33.8%	6/30 40.0%	15/34 31.7%	3/17 23.3%	9/19 16.7%
Terminal 98	25	20	24	19	14	10

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.

Table 16 below provides the significance levels of the tests of homogeneity and trend over dose groups as proposed in Section 1.3.1.1, above.

Table 16. Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1&2, 3-6	0.3264	0.2464	0.0033	0.0045
Trend over Groups 1&2, 3-6	0.0429	0.0429	0.0001	0.0005
Homogeneity over Groups 1&2 vs 6	0.0404	0.0260	0.0006	0.0006

Unlike the situation with rats, in mice there is evidence of a dose related decrement to survival. From figures A.1.3 and A.1.4 in Appendix 1, in mice there is a general decrease in survival over doses. Overall, in male mice the highest mortality is in the high dose group with the lowest mortality in control group 1. Interestingly, the survival curve of the second control group in male mice is generally fairly closely intertwined with the remaining dose groups, with survival generally between that of the high dose group and the first control group. This explains the possibly inconsistent results of the statistical tests, i.e. no evidence of an overall lack of homogeneity (LR p = 0.3264, Wilcoxon p = 0.2464), but some slightly equivocal evidence of a trend (LR and Wilcoxon p = 0.0429) and a statistically significant difference between the high dose group and pooled controls (LR p = 0.0404, Wilcoxon p = 0.0260). Results for female mice are much stronger. In Figure A.1.4, the high dose group in female mice also has the highest mortality, closely followed by the medium high dose group. The remaining dose groups are generally intertwined. This explains the statistically significant results when testing for overall lack of homogeneity, trend, and no pairwise difference between the high dose group and pooled controls (all six p ≤ 0.0045).

Tumorigenicity analysis:

As discussed in Section 1.3.1.5, the Haseman-Lin-Rahman rules are used to adjust for the multiplicity of tests in the carcinogenicity analysis:

Table 17. Potentially Statistically Significant Neoplasms in Mice (All Female)

	Veh 1	Veh 2	Low	Med	Hi	Trend	High vs Veh	Med-Hi vs Veh	Medium vs Veh	Low vs Veh
ADRENAL, MEDULLA										
B-PHEOCHROMOCYTOMA	0	0	0	0	3	0	0.3776	.	0.0261	.
HARDERIAN GLAND										
Adenoma/Carcinoma	5	3	7	3	5	8	0.0354	0.0489	0.3406	0.5383	0.1383
MAMMARY, FEMALE										
Adenoma/Carc./-sarcoma/-canth.	2	1	13	19	26	20	0.0000	0.0000	0.0000	0.0000	0.0000
M-ADENOACANTHOMA	1	0	7	6	7	5	0.1028	0.0080	0.0011	0.0061	0.0014
M-CARCINOMA	2	1	7	12	18	13	0.0011	0.0000	0.0000	0.0002	0.0113
M-CARCINOSARCOMA	0	0	0	1	2	2	0.0259	0.0800	0.0898	0.3333	.
OVARY										
Cystad./Gran./Thecal/Tubul.	1	0	2	4	1	2	0.2543	0.1961	0.5073	0.0446	0.2307
PANCREAS										
B-ISLET CELL ADENOMA	0	0	0	1	3	1	0.1546	0.2857	0.0261	0.3333	.
PITUITARY										
B-ADENOMA, PARS DISTALIS	3	4	11	17	27	29	0.0000	0.0000	0.0000	0.0001	0.0068

In male mice no tests of trend or tests of pairwise differences from the control achieved the nominal 0.05 level of significance, let alone the levels adjusted for multiplicity. However, in female mice there is some strong evidence a carcinogenic response in mammary and

pituitary tumors. In particular, the tests of overall trend and pairwise comparison between the high dose group and control in mammary malignant carcinoma were statistically significant ($p = 0.0011 < 0.005$ and $p < 0.00005 < 0.01$, respectively), as were the tests of pooled tumors ($p < 0.00005 < 0.005$ and $p < 0.00005 < 0.01$). Similarly the tests of overall trend and pairwise comparison between the high dose group and control in pituitary pars distalis adenoma were highly statistically significant ($p = 0.00005 < 0.005$ and 0.01, respectively). Note the relative constancy of mammary adenoacanthoma across the actual Lurasidone treatment groups implies there is no strong evidence of trend, the comparison the high dose to control is statistically significant ($p = 0.008 < 0.05$). Again, incorporating the other pairwise comparisons can be expected to raise rhe nominal type I error rate to something above the rough 10% level. But if one accepts that potential inflation, the pairwise comparisons of the medium-high and medium dose groups in mammary malignant carcinoma were statistically significant ($p < 0.00005 < 0.01$ and $p = 0.0002 < 0.01$, respectively), while comparison in the low dose group was close to adjusted statistical significance ($p = 0.0113 \approx 0.01$). However, in pooled mammary tumors all these comparisons would also be statistically significant (all three $p < 0.00005 < 0.01$). These pairwise comparisons in mammary adenocanthoma would also be labeled as statistically significant (all $p = 0.0011, 0.0061, 0.0014 < 0.05$). Similarly the pairwise comparisons of the medium-high, medium, and low dose groups to the pooled controls in pituitary pars distalis adenoma would also be classified as statistically significant (all three $p < 0.00005$, $p = 0.0001$, $p = 0.0068 < 0.01$). The comparison between the medium dose group and pooled vehicle in pooled cancers of the ovary would classified as statistically significant ($p = 0.0446 < 0.05$), though only barely. Finally, the comparisons of the medium-high dose group to the the pooled vehicle in adrenal pheochromocytoma and islet cell adenoma of the pancrease were statistically significant ($p = 0.0261 < 0.05$). Following the Haseman-Lin-Rahman rules above, no other tests achieved statistical significance, though several were close.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see Section 1.1 above.

APPENDICES:

Appendix 1. Survival Analysis

Simple summary life tables in mortality are presented in the report (Tables 9, 10, 14, and 15 above). Kaplan-Meier estimated survival curves across dose groups for each gender in each study are displayed in Figures A.1.1-A.1.4 below. These plots include 95% confidence intervals around each survival curve (colored area around each curve). Contrary to the sponsor's analysis for all tests the two control groups are pooled (please see Section 1.3.1.1 for a discussion). The plots are also supported by tests of homogeneity in survival over the different treatment groups and the pooled controls, tests of trend in survival over increasing dose, and the results of pairwise comparisons between the high dose group and pooled controls in Table A.1.1. below. One might note that the log rank tests places greater weight on later events, while the Wilcoxon test tends to weight weights them more equally, and thus places less weight on earlier events than does the log rank test.

Table A.1.1 Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1&2, 3-5	0.0243	0.0204	0.4451	0.7525
Trend over Groups 1&2, 3-5	0.0131	0.0060	0.2749	0.5144
Difference between Groups 1&2 vs 5	0.0332	0.0282	0.1639	0.4052
Mice Homogeneity over Groups 1&2, 3-6	0.3264	0.2464	0.0033	0.0045
Trend over Groups 1&2, 3-6	0.0429	0.0429	0.0001	0.0005
Difference between Groups 1&2 vs 6	0.0404	0.0260	0.0006	0.0006

From Figure A.1.2 below, in female rats it seems that the the dose groups are all quite intertwined, consistent with the hypothesis of no heterogeneity, trend, or differences between the high dose and pooled controls (all six $p \geq 0.1639$). In Figure A.1.1 for male rats there does seem to be evidence of dose related differences, but after a starting period, the high dose group has the lowest mortality (i.e., greatest survival), clearly separated from the remaining groups. Meanwhile, the low dose group generally has the highest mortality, although the survival curves of the remaining dose groups are generally closely intertwined with the survival curve of this group. This is consistent with the results of the various statistically significant tests in male rats (Overall homogeneity: LR p = 0.0243, Wilcoxon p = 0.0204, Trend: LR p = 0.0131, Wilcoxon p = 0.0006, High vs. control: LR p = 0.0332, Wilcoxon p = 0.0282).

Figure A.1.1 Kaplan-Meier Survival Curves for Male Rats

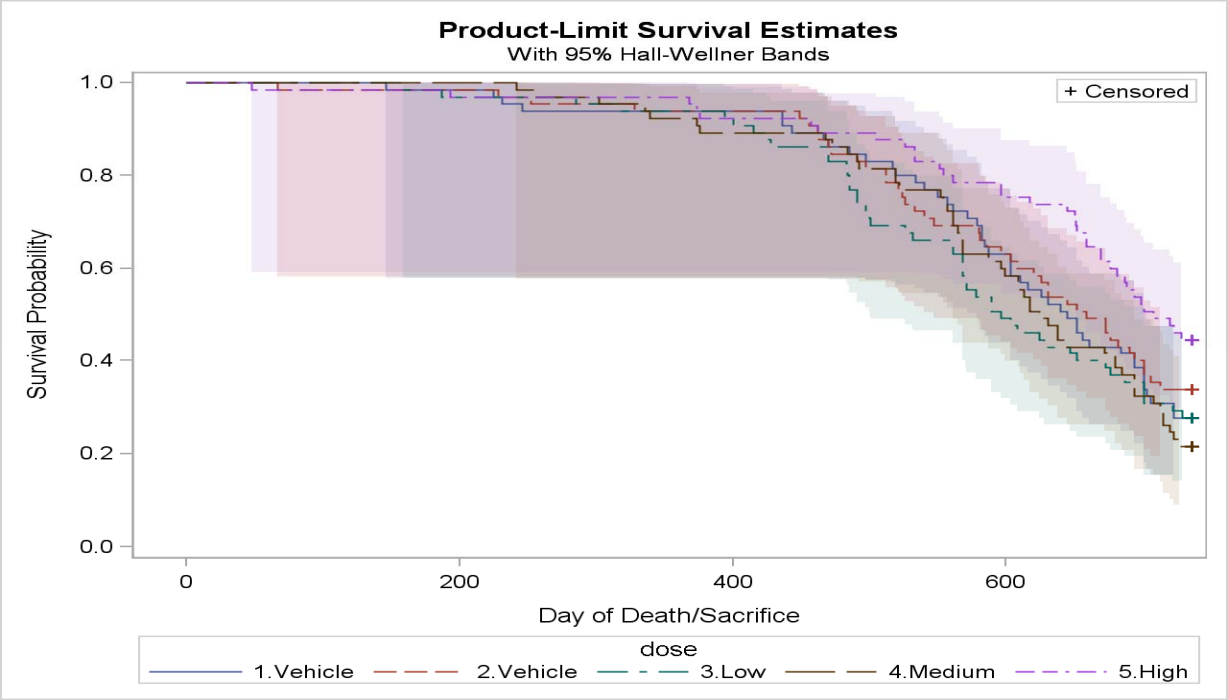
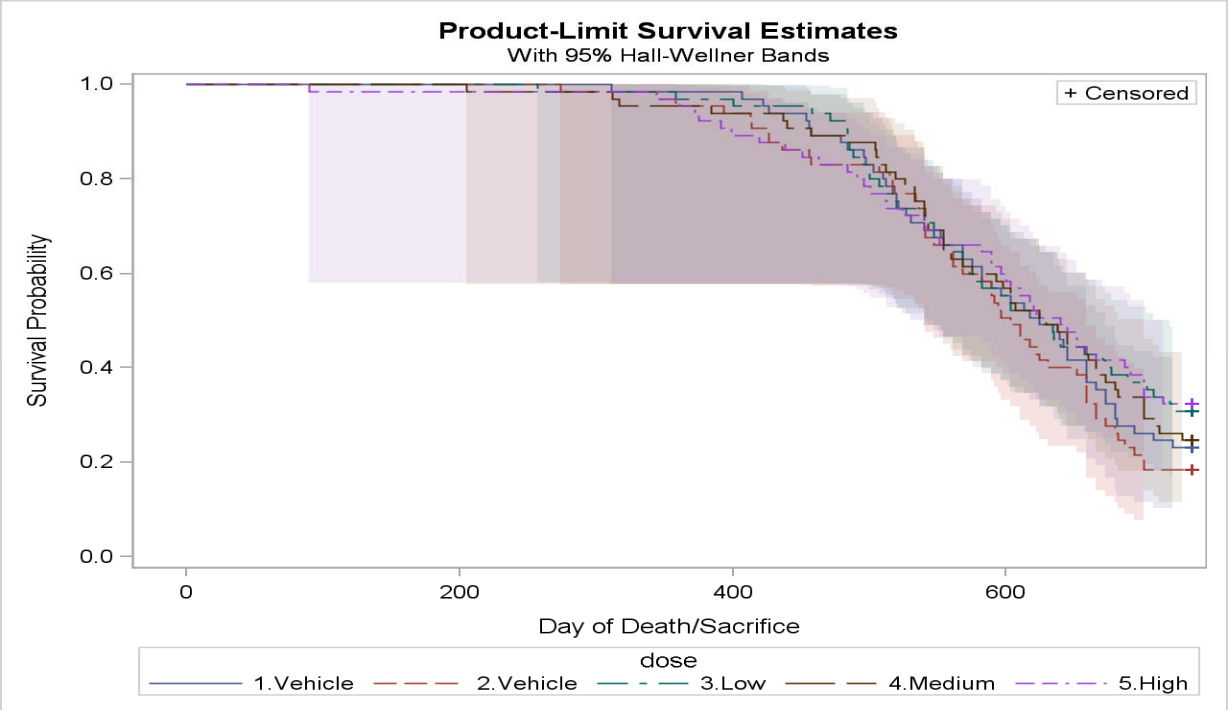


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



Unlike the situation with rats, in mice there is evidence of a dose related decrement to survival. From Figure A.1.3 below, in male mice there is a general decrease in survival over doses. Overall, the highest mortality is in the high dose group with the lowest mortality in control group 1. Interestingly, the survival curve of the second control group in male mice is generally fairly closely intertwined with the remaining dose groups, with survival generally between the high dose group and the first control group. This explains the possibly inconsistent results of the statistical tests, i.e. no evidence of an overall lack of homogeneity (LR p = 0.3264, Wilcoxon p = 0.2464), but some slightly equivocal evidence of a trend (LR and Wilcoxon p = 0.0429) and difference between the high dose group and pooled controls (LR p = 0.0404, Wilcoxon p = 0.0260). Results for female mice are much stronger. In Figure A.1.4, the high dose group in female mice also has the highest mortality, closely followed by the medium high dose group. The remaining dose groups are generally intertwined. This explains the statistically significant results when testing for overall lack of homogeneity, trend, and no pairwise difference between the high dose group and pooled controls (all six p ≤ 0.0045).

Figure A.1.3 Kaplan-Meier Survival Curves for Male Mice

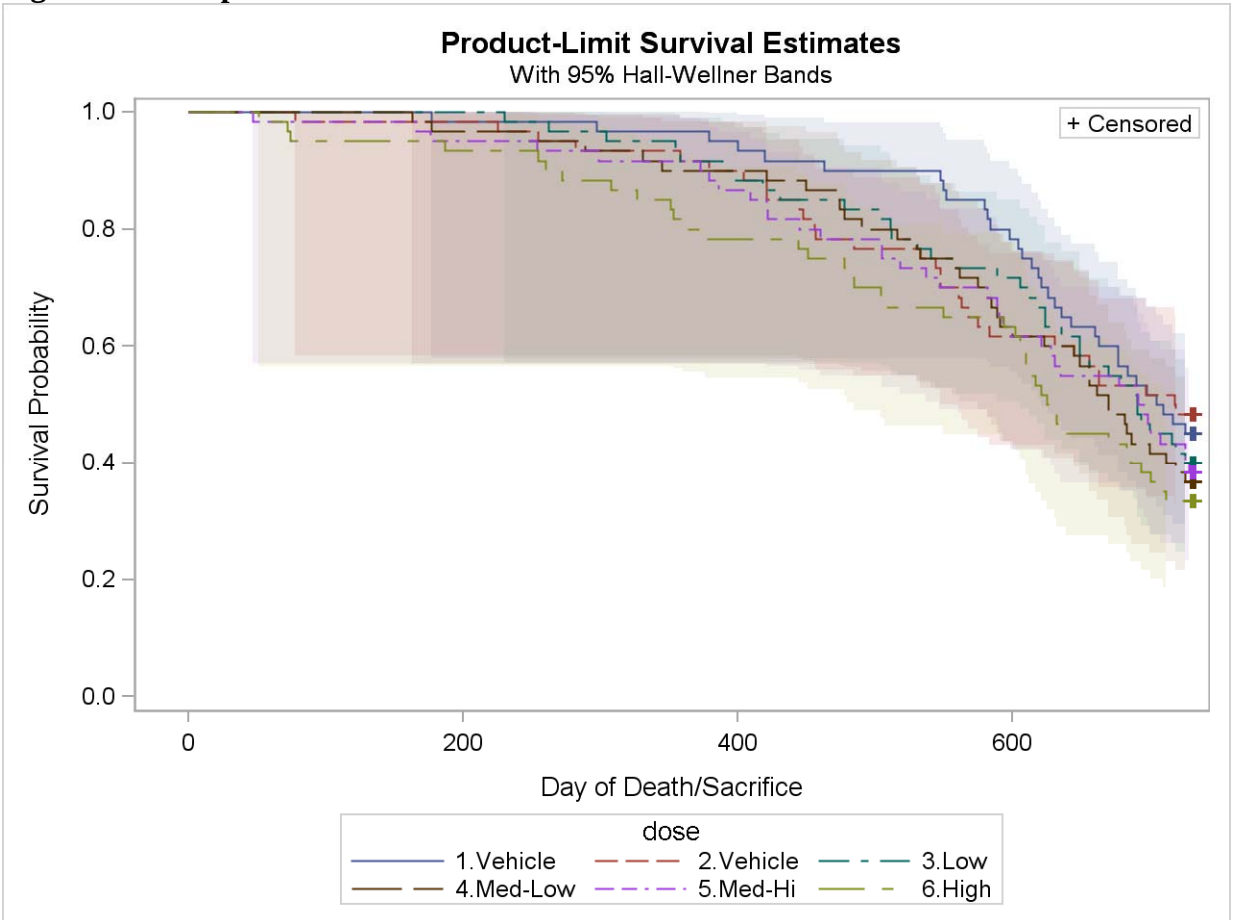
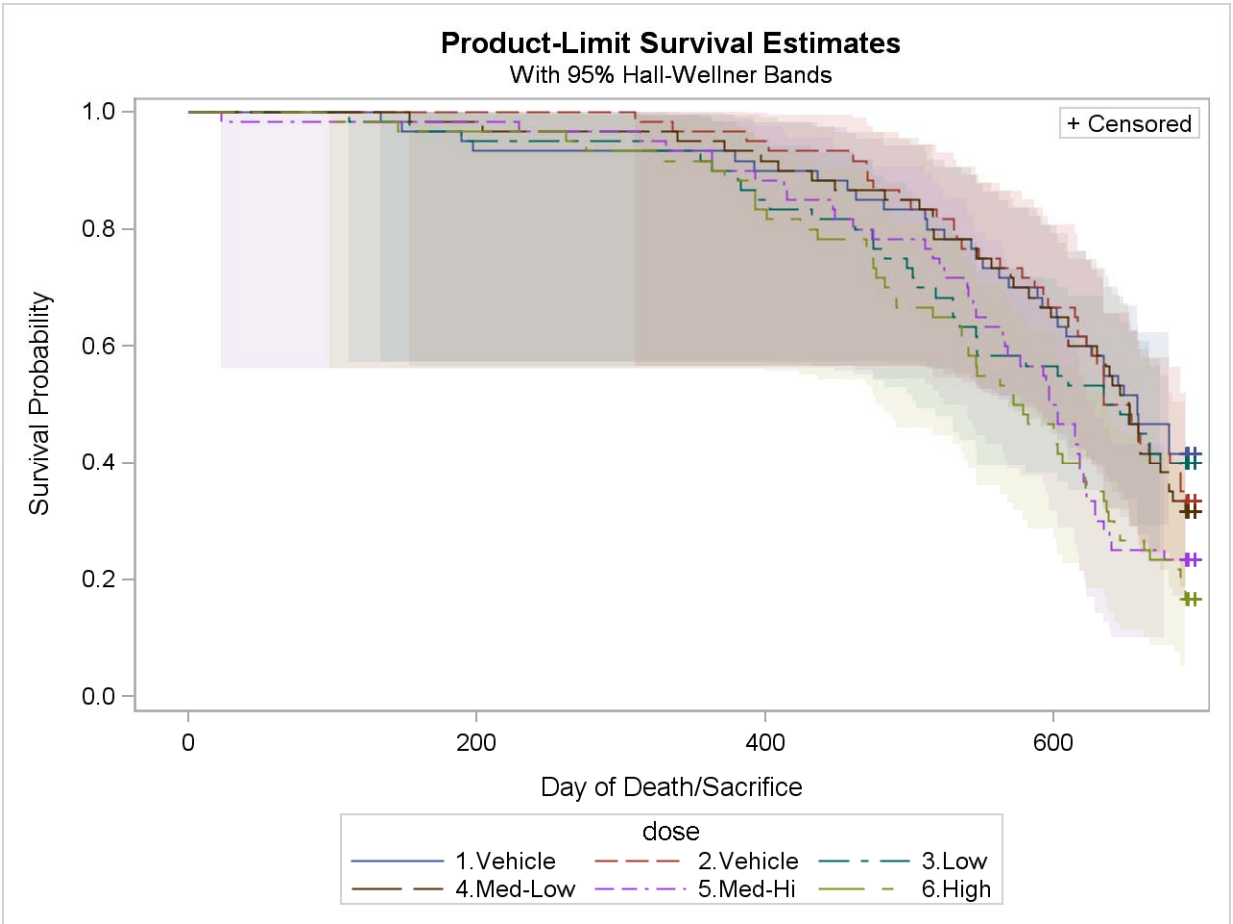


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



Appendix 2. FDA Poly-k Tumorigenicity Analysis

Tables A.2.1 through A.2.6 given below, display, in each study, for each gender, the organ by tumor combination, the number of animals with one or more of the specified tumor in each treatment group, plus the statistical significance levels of the tests of no trend or no pairwise difference between the specified treatment groups and the pooled vehicle. The tumor incidences in each vehicle control are listed separately, but the results of statistical tests are based on pooled controls.

The poly-k test, here with k=3, modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The tests used here are small sample exact permutation tests of tumor incidence. These do assume all marginal totals are fixed, a debatable assumption. This assumption implies that in the pairwise tests when one dose group has no tumors of the specific type and the other does, there is only one permutation of this pattern. Since that means that the only permutation of the data is the one observed, that means that all possible permutations are as extreme as the pattern observed, and thus the significance level of the observed pattern can be logically expressed as 1.0. One could use the same sort of argument when there were no tumors of the specific type being analyzed in either cell of the 2x2 table corresponding to a pairwise comparison. Then an argument could be made that the p-value for this test should also be 1.0. However, largely for readability, in the tables below these p-values are considered as missing (i.e., corresponding to a null test), denoted by ".". Note that StatXact adjusts for the variance, which would be 0. Then the significance levels of the test statistics are based on the result of a division by 0, i.e., undefined, and hence StatXact codes these p-values as missing.

Up until recently, the Division has usually emphasized so-called Peto carcinogenicity tests, which require accurate specification of cause of death. This is the testing methodology used by the Sponsor. It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that the poly-k modification of the Cochran-Armitage tests of trend has been recommended over such Peto tests.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.5 are usually applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors (incidence > 1%) and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors.

Tables A.2.1 and A.2.2 in rats and mice, respectively, show the tumors that had at least one mortality adjusted test whose nominal statistical significance was at least 0.05. Tables A.2.3 and A.2.4 tediously display all test results for male and female rats, respectively, while Tables A.2.5 and A.2.6 present similar results in male and female mice. The p-values of the

poly-k test are based on exact tests from StatXact as discussed above. As also noted above, the period '.' denotes the p-values of tests of dose groups with no tumors in any group.

In Table A.2.1, in female rats, following the adjustment for multiplicity to get an overall rough 10% error rate and using the incidence in the no treatment group to decide if a tumor is rare or not, we would conclude, that the test of trend in malignant mammary carcinoma was statistically significant ($p = 0.0008 < 0.005$), as was the test comparing the high dose to the pooled vehicle ($p = 0.0009 < 0.01$). No other tests of trend or comparisons between the high dose and controls achieved the multiplicity adjusted significance levels. Applying the rule for pairwise comparisons to other groups than the high dose group can be expected to inflate the type I error rate above the rough 10% level established in Lin and Rahman (2007). But if such a rule is used, the comparison beween the low dose and vehicle in hemato neoplasia lymphoma in male rats would be statistically significant, though barely ($p = 0.0396 < 0.05$), as would be the comparison of pooled adenoma and other neoplasms in the mammarys of male rats between vehicle and medium dose groups ($p = 0.0372 < 0.05$). In male rats the tests of differences between the medium dose group and pooled controls in skin fibroma/fibrosarcoma would also be statistically significant ($p = 0.0064 < 0.01$) as would be the test of malignant fibrosarcoma ($p = 0.0156 < 0.05$). Again accepting the probable inflation of overall Type I error, we would also note that the pairwise comparison between the medium dose group and pooled controls in mammary adenoma and carcinoma in female rats would also be classified as being statistically significant ($p = 0.0044, 0.0027 < 0.01$, respectively). No other comparisons would meet these looser criteria.

Table A.2.1 Potentially Statistically Significant Neoplasms in Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low ve Veh
N	65	65	65	65	65				
Male Rats									
HEMATO NEOPLASIA									
M-LYMPHOMA	0	1	4	1	2	0.4121	0.3047	0.5432	0.0396
MAMMARY, MALE									
Adenoma/Carc./Fibro.	0	1	1	4	1	0.5059	0.6019	0.0372	0.5284
SKIN									
Fibroma/Fibrosarcoma	2	0	2	7	1	0.6429	0.2982	0.0064	0.3789
M-FIBROSARCOMA	1	0	2	5	0	0.8026	0.3630	0.0156	0.2364
Female Rats									
MAMMARY, FEMALE									
Adenoma/Carc./Fibro./mixed	38	38	41	50	42	0.1771	0.1966	0.0121	0.3523
B-ADENOMA	11	5	10	20	12	0.2538	0.2477	0.0044	0.4292
M-CARCINOMA	19	14	21	30	32	0.0008	0.0009	0.0027	0.2483
THYROID									
Adenoma/Carc. C cell	4	4	2	4	8	0.0398	0.1423	0.3807	0.7363

In male mice no tests of trend or tests of pairwise differences from the control achieved the nominal 0.05 level of significance, let alone the levels adjusted for multiplicity. However, in female mice there is some strong evidence of carcinogenicity in mammary and pituitary tumors. In particular, the tests of overall trend and pairwise comparison between the high dose

group and control in mammary malignant carcinoma was statistically significant ($p = 0.0011 < 0.005$ and $p < 0.00005 < 0.01$, respectively), as were the tests of pooled tumors ($p < 0.00005 < 0.005$ and $p < 0.00005 < 0.01$). Similarly the tests of overall trend and pairwise comparison between the high dose group and control in pituitary pars distalis adenoma was highly statistically significant ($p = 0.00005 < 0.005$ and $p < 0.00005 < 0.01$, respectively). Note the relative constancy of mammary adenoacanthoma across the actual Lurasidone treatment groups implies there is no strong evidence of trend, the comparison the high dose to control is statistically significant ($p = 0.008 < 0.05$). Again, incorporating the other pairwise comparisons can be expected to raise rhe nominal type I error rate to something above the rough 10% level. But if one accepts that potential inflation, the pairwise comparisons of the medium-high and medium dose groups in mammary malignant carcinoma were statistically significant ($p < 0.00005 < 0.01$ and $p = 0.0002 < 0.01$, respectively), while comparison in the low dose group was close to adjusted statistical significance ($p = 0.0113 \approx 0.01$). However, in pooled mammary tumors all these comparisons would also be statistically significant (all three $p < 0.00005 < 0.01$). These pairwise comparisons in mammary adenocanthoma would also be labeled as statistically significant (all $p = 0.0011, 0.0061, 0.0014 < 0.05$). Similarly the pairwise comparisons of the medium-high, medium, and low dose groups to the pooled controls in pituitary pars distalis adenoma would also be classified as statistically significant (all three $p < 0.00005$, $p = 0.0001$, $p = 0.0068 < 0.01$). The comparison between the medium dose group and pooled vehicle in pooled cancers of the ovary would classified as statistically significant ($p = 0.0446 < 0.05$), though only barely. Finally, the comparisons of the medium-high dose group to the the pooled vehicle in adrenal pheochromocytoma and islet cell adenoma of the pancrease were statistically significant ($p = 0.0261 < 0.05$). Following the Haseman-Lin-Rahman rules above, no other tests achieved statistical significance, though several were close.

Table A.2.2 Potentially Statistically Significant Neoplasms in Mice (All Female)

	Veh 1	Veh 2	Low	Med	Hi	Trend	High vs Veh	Med-Hi vs Veh	Medium vs Veh	Low vs Veh
ADRENAL, MEDULLA										
B-PHEOCHROMOCYTOMA	0	0	0	0	3	0 0.3776	.	0.0261	.	.
HARDERIAN GLAND										
Adenoma/Carcinoma	5	3	7	3	5	8 0.0354	0.0489	0.3406	0.5383	0.1383
MAMMARY, FEMALE										
Adenoma/Carc./-sarcoma/-canth.	2	1	13	19	26	20 0.0000	0.0000	0.0000	0.0000	0.0000
M-ADENOACANTHOMA	1	0	7	6	7	5 0.1028	0.0080	0.0011	0.0061	0.0014
M-CARCINOMA	2	1	7	12	18	13 0.0011	0.0000	0.0000	0.0002	0.0113
M-CARCINOSARCOMA	0	0	0	1	2	2 0.0259	0.0800	0.0898	0.3333	.
OVARY										
Cystad./Gran./Thecal/Tubul.	1	0	2	4	1	2 0.2543	0.1961	0.5073	0.0446	0.2307
PANCREAS										
B-ISLET CELL ADENOMA	0	0	0	1	3	1 0.1546	0.2857	0.0261	0.3333	.
PITUITARY										
B-ADENOMA, PARS DISTALIS	3	4	11	17	27	29 0.0000	0.0000	0.0000	0.0001	0.0068

Complete incidence tables are presented below:

Table A.2.3 Incidence and Statistical Tests of Carcinogenicity in Male Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low vs Veh
ADIPOSE TISSUE									
B-HIBERNOMA	0	1	1	0	0	0.6944	0.3630	0.3228	0.5359
B-LIPOMA	1	0	1	0	0	0.6953	0.3630	0.3228	0.5284
ADRENAL, CORTEX									
B-ADENOMA	1	1	3	0	2	0.4397	0.4606	0.5432	0.1827
ADRENAL, MEDULLA									
B-PHEOCHROMOCYTOMA	5	3	8	4	9	0.1405	0.1050	0.5838	0.0776
M-MALIGNANT PHEOCHROMOCYTOMA	0	0	0	0	1	0.2315	0.3676	.	.
Pheochromocytoma all	5	3	8	4	10	0.0894	0.0664	0.5838	0.0776
AUDITORY SEB GL									
M-CARCINOMA, SQUAMOUS CELL	0	0	0	0	1	0.2315	0.3676	.	.
BRAIN									
M-ASTROCYTOMA	0	1	3	2	1	0.5584	0.6019	0.2505	0.0985
M-GRANULAR CELL TUMOR	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
CALVARIUM/SKULL									
M-SARCOMA	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
M-SARCOMA, OSTEOGENIC	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
CAVITY, ABDOM									
B-LIPOMA	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
EPIDIDYMIS									
M-MALIGNANT MESOTHELIOMA	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
FOOT/FOOTPAD									
B-PAPILLOMA, SQUAMOUS	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
HEMATO NEOPLASIA									
M-LEUKEMIA, MYELOID	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
M-LYMPHOMA	0	1	4	1	2	0.4121	0.3047	0.5432	0.0396
M-SARCOMA, HISTIOCYTIC	1	3	4	0	3	0.4749	0.5098	0.7945	0.2207
KIDNEY									
Adenoma/Carc. Tubular	0	0	0	2	0	0.4047	.	0.1025	.
B-ADENOMA, TUBULAR CELL	0	0	0	1	0	0.2279	.	0.3228	.
B-LIPOMA	0	0	0	0	1	0.2315	0.3676	.	.
Lipoma/Liposarcoma	0	1	0	1	1	0.3141	0.6019	0.5432	0.3120
M-CARCINOMA, TUBULAR CELL	0	0	0	1	0	0.2279	.	0.3228	.
M-LIPOSARCOMA	0	1	0	1	0	0.5098	0.3630	0.5432	0.3120
LIVER									
Adenoma/Carc. Hepatocell	3	3	1	2	2	0.6513	0.6075	0.5084	0.7027
B-ADENOMA, HEPATOCELLULAR	2	3	1	1	1	0.8033	0.7102	0.6328	0.6102
M-CARCINOMA, HEPATOCELLULAR	1	0	0	1	1	0.3091	0.5959	0.5432	0.3120
LN, MESENTERIC									
B-HEMANGIOMA	0	1	1	0	0	0.6953	0.3630	0.3228	0.5284
M-ANGIOSARCOMA	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
M-HEMANGIOSARCOMA	2	0	1	0	0	0.8501	0.5959	0.5432	0.6779
LUNG									
B-ADENOMA, BRONCHIOLAR-ALVEO	2	0	0	0	0	0.8411	0.5959	0.5432	0.5284
MAMMARY, MALE									
Adenoma/Carc./Fibro.	0	1	1	4	1	0.5059	0.6019	0.0372	0.5284
B-ADENOMA	0	1	0	1	0	0.5098	0.3630	0.5432	0.3120
B-FIBROADENOMA	0	0	0	1	0	0.2279	.	0.3228	.
M-CARCINOMA	0	0	1	2	1	0.2539	0.3676	0.1025	0.3120
MUSCLE, SKELETAL									
M-FIBROSARCOMA	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
M-SARCOMA	0	0	0	1	0	0.2279	.	0.3228	.
PANCREAS									
Adenoma/Carc. Islet Cells	7	10	10	9	14	0.1918	0.2080	0.4877	0.3189
B-ADENOMA, ACINAR CELL	2	1	0	1	0	0.8396	0.7448	0.3885	0.6779
B-ADENOMA, ISLET CELL	6	8	9	8	11	0.3150	0.3049	0.4150	0.2538
M-CARCINOMA, ISLET CELL	1	2	1	1	3	0.2107	0.3876	0.3885	0.3789

Table A.2.3 (cont.) Incidence and Statistical Tests of Carcinogenicity in Male Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low vs Veh
PARATHYROID									
B-ADENOMA	1	1	0	0	0	0.8411	0.5959	0.5432	0.5284
PINNA									
B-PAPILLOMA, SQUAMOUS CELL	0	0	0	1	0	0.2279	.	0.3228	.
PITUITARY									
Adenoma/Carcinoma	35	36	34	40	35	0.7533	0.6558	0.2870	0.5255
B-ADENOMA	34	36	33	40	35	0.6991	0.6103	0.2475	0.4327
M-CARCINOMA	1	0	1	0	0	0.6953	0.3630	0.3228	0.5284
SKIN									
B-ADENOMA, BASAL CELL	1	1	2	0	1	0.6113	0.2982	0.5432	0.3690
B-ADENOMA, SEBACEOUS GLAND	0	0	0	1	0	0.2279	.	0.3228	.
B-FIBROMA	1	0	0	2	1	0.2879	0.5959	0.2435	0.3120
B-KERATOACANTHOMA	2	1	3	0	1	0.7872	0.4606	0.6930	0.2747
B-LIPOMA	1	0	0	2	0	0.5677	0.3630	0.2435	0.3120
B-PAPILLOMA, SQUAMOUS CELL	2	2	0	0	0	0.9755	0.8395	0.7945	0.7809
Fibroma/Fibrosarcoma	2	0	2	7	1	0.6429	0.2982	0.0064	0.3789
M-CARCINOMA	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
M-CARCINOMA, SQUAMOUS CELL	1	0	0	1	0	0.5098	0.3630	0.5432	0.3120
M-FIBROSARCOMA	1	0	2	5	0	0.8026	0.3630	0.0156	0.2364
M-LIPOSARCOMA	1	0	0	0	0	0.6000	0.3630	0.3228	0.3120
M-NEUROFIBROSARCOMA	0	0	1	0	1	0.2250	0.3676	.	0.3120
SPINAL CORD									
M-ASTROCYTOMA	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
SPLEEN									
M-HEMANGIOSARCOMA	0	1	0	0	1	0.4102	0.6019	0.3228	0.3120
Systemic									
B-HEMANGIOMA	0	1	1	0	1	0.4377	0.6019	0.3228	0.5284
Hemangioma/-sarcoma	0	1	1	0	1	0.4377	0.6019	0.3228	0.5284
TAIL									
B-PAPILLOMA, SQUAMOUS CELL	0	1	1	1	0	0.6295	0.3630	0.5432	0.5284
TESTIS									
B-INTERSTITIAL CELL TUMOR	1	0	0	1	0	0.5098	0.3630	0.5432	0.3120
THORACIC CAVITY									
B-HIBERNOMA	1	0	0	1	0	0.5107	0.3630	0.5503	0.3120
THYMUS									
M-HIBERNOMA	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
M-THYMIC CARCINOMA	0	1	0	0	0	0.6000	0.3630	0.3228	0.3120
THYROID									
Adenoma/Carc. Foll. cell	5	2	3	1	3	0.6475	0.5369	0.7961	0.3804
B-ADENOMA, C-CELL	8	7	5	11	9	0.4004	0.5307	0.1622	0.6241
B-ADENOMA, FOLLICULAR CELL	4	2	2	1	2	0.7202	0.6186	0.7241	0.4805
M-CARCINOMA, FOLLICULAR CELL	1	0	1	0	1	0.4320	0.5959	0.3228	0.5284
URINARY BLADDER									
B-HEMANGIOMA	0	0	0	0	1	0.2315	0.3676	.	.
M-CARCINOMA, TRANSITIONAL CE	0	0	0	1	0	0.2279	.	0.3228	.

Table A.2.4 Incidence and Statistical Tests of Carcinogenicity in Female Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low vs Veh
ADRENAL, CORTEX									
Adenoma/Carcinoma	1	1	0	1	1	0.4285	0.2725	0.2725	0.5794
B-ADENOMA	1	0	0	0	1	0.3655	0.5723	0.3445	0.3500
M-CARCINOMA	0	1	0	1	0	0.4908	0.3445	0.5723	0.3500
ADRENAL, MEDULLA									
B-GANGLIONEUROMA	0	0	0	1	0	0.2030	.	0.3445	.
B-PHEOCHROMOCYTOMA	1	2	4	3	1	0.7535	0.4322	0.3448	0.1887
Ganglioneuroma/Neurofibro.	0	0	1	1	0	0.4059	.	0.3445	0.3500
M-MALIGNANT PHEOCHROMOCYTOMA	0	0	0	1	0	0.2030	.	0.3445	.
M-NEUROFIBROSARCOMA	0	0	1	0	0	0.4059	.	.	0.3500
Pheochromocytoma all	1	2	4	4	1	0.7642	0.4322	0.1887	0.1887
BRAIN									
M-ASTROCYTOMA	1	1	1	2	0	0.7377	0.5685	0.4322	0.2761
M-OLIGODENDROGLIOMA	0	0	0	0	1	0.2069	0.3500	.	.
CAVITY, ABDOM									
B-HEMANGIOMA	0	0	1	0	0	0.4059	.	.	0.3500
B-HIBERNOMA	0	1	0	0	0	0.6108	0.3417	0.3417	0.3471
B-LIPOMA	0	2	0	0	0	0.8521	0.5723	0.5723	0.5794
CERVIX									
B-FIBROMA	0	0	0	1	0	0.2030	.	0.3445	.
B-GRANULAR CELL TUMOR	0	1	0	0	0	0.6139	0.3445	0.3445	0.3500
B-LEIOMYOMA	0	1	0	0	0	0.6139	0.3445	0.3445	0.3500
B-POLYP, ENDOMETRIAL STROMAL	1	0	0	0	0	0.6139	0.3445	0.3445	0.3500
M-SARCOMA, ENDOMETRIAL STROM	2	1	0	0	0	0.9411	0.7147	0.7147	0.7217
CLITORAL GLAND									
B-PAPILLOMA, DUCTAL	0	0	0	0	1	0.2030	0.3445	.	.
M-CARCINOMA, SQUAMOUS CELL	0	0	1	0	0	0.4059	.	.	0.3500
FOOT/FOOTPAD									
B-SQUAMOUS PAPILLOMA	0	0	0	1	0	0.2020	.	0.3500	.
GINGIVA									
M-CARCINOMA, SQUAMOUS CELL	0	0	0	1	0	0.2030	.	0.3445	.
HEMATO NEOPLASIA									
M-LYMPHOMA	0	0	1	2	1	0.2170	0.3500	0.1206	0.3500
M-SARCOMA, HISTIOCYTIC	1	1	0	1	1	0.4357	0.2800	0.2725	0.5794
ILEUM									
M-LEIOMYOSARCOMA	0	1	0	0	0	0.6139	0.3445	0.3445	0.3500
LIVER									
B-ADENOMA, HEPATOCELLULAR	2	2	0	1	0	0.9051	0.8171	0.5585	0.8231
B-CHOLANGIOMA	1	0	0	1	0	0.4908	0.3445	0.5723	0.3500
LN, MESENTERIC									
M-HEMANGIOSARCOMA	0	0	1	0	0	0.4059	.	.	0.3500
LUNG									
B-ADENOMA, BRONCHIOLAR-ALVEO	0	1	0	0	0	0.6139	0.3445	0.3445	0.3500
MAMMARY, FEMALE									
Adenoma/Carc./Fibro./mixed	38	38	41	50	42	0.1771	0.1966	0.0121	0.3523
B-ADENOMA	11	5	10	20	12	0.2538	0.2477	0.0044	0.4292
B-FIBROADENOMA	24	28	26	22	23	0.7272	0.6603	0.7435	0.5720
M-CARCINOMA	19	14	21	30	32	0.0008	0.0009	0.0027	0.2483
M-MIXED MAMMARY TUMOR	1	0	0	0	0	0.6139	0.3445	0.3445	0.3500
M-SQUAMOUS CELL CARCINOMA	0	1	0	1	0	0.4920	0.3445	0.5794	0.3500
MUSCLE, SKELETAL									
M-FIBROSARCOMA	0	0	0	0	1	0.2069	0.3500	.	.
M-RHABOMYOSARCOMA	1	0	0	0	0	0.6139	0.3445	0.3445	0.3500
PANCREAS									
Adenoma/Carc. Islet Cells	5	2	3	5	5	0.2567	0.4058	0.4058	0.4797
B-ADENOMA, ISLET CELL	3	2	1	3	4	0.1632	0.3747	0.5634	0.6869
M-CARCINOMA, ISLET CELL	2	0	2	2	1	0.5209	0.2800	0.4374	0.4374

Table A.2.4 (cont.) Incidence and Statistical Tests of Carcinogenicity in Female Rats

	Incidence					Significance Levels			
	Veh 1	Veh 2	Low	Med	High	Trend	High vs Veh	Medium vs Veh	Low vs Veh
PINNA									
B-NEUROLEMMOMA	1	1	0	0	0	0.8497	0.5685	0.5685	0.5756
PITUITARY									
B-ADENOMA	54	58	55	40	48	0.8563	0.8764	0.9993	0.7295
M-CARCINOMA	2	0	0	3	2	0.1869	0.4374	0.2226	0.5794
SKIN									
B-ADENOMA, BASAL CELL	1	1	0	0	0	0.8521	0.5723	0.5723	0.5794
B-FIBROMA	1	0	0	1	1	0.2848	0.5794	0.5723	0.3500
B-LIPOMA	1	1	0	0	0	0.8521	0.5723	0.5723	0.5794
B-PAPILLOMA, SQUAMOUS CELL	0	1	0	0	0	0.6139	0.3445	0.3445	0.3500
M-FIBROSARCOMA	1	1	4	1	0	0.9070	0.5685	0.2686	0.1086
M-NEUROFIBROSARCOMA	0	1	0	0	0	0.6139	0.3445	0.3445	0.3500
Systemic									
B-HEMANGIOMA	0	0	1	0	0	0.4059	.	.	0.3500
Hemangioma/-sarcoma	0	0	1	0	0	0.4059	.	.	0.3500
THYROID									
Adenoma/Carc. C cell	4	4	2	4	8	0.0398	0.1423	0.3807	0.7363
Adenoma/Carc. Foll. cell	0	2	1	2	0	0.7400	0.5723	0.4276	0.2800
B-ADENOMA, C-CELL	4	4	2	3	7	0.0767	0.2311	0.5480	0.7363
B-ADENOMA, FOLLICULAR CELL	0	1	1	1	0	0.6023	0.3445	0.5723	0.5794
M-CARCINOMA, C-CELL	0	0	0	1	1	0.1232	0.3445	0.3445	.
M-CARCINOMA, FOLLICULAR CELL	0	1	0	1	0	0.4908	0.3445	0.5723	0.3500
TONGUE									
M-CARCINOMA, SQUAMOUS CELL	0	0	0	1	0	0.2030	.	0.3445	.
UTERUS									
B-LEIOMYOMA	0	0	1	0	1	0.2080	0.3445	.	0.3500
B-POLYP, ENDOMETRIAL, STROMA	3	2	1	3	0	0.9153	0.8791	0.5626	0.6759
Leyomyoma/Carc./Polyp	3	2	2	3	2	0.5463	0.4424	0.5626	0.4551
M-CARCINOMA	0	0	0	0	1	0.2030	0.3445	.	.
VAGINA									
All tumors	1	0	0	1	2	0.0969	0.2725	0.5723	0.3500
B-FIBROMA	0	0	0	0	1	0.2030	0.3445	.	.
B-GRANULAR CELL TUMOR	1	0	0	0	1	0.3655	0.5723	0.3445	0.3500
M-LEIOMYOSARCOMA	0	0	0	1	0	0.2030	.	0.3445	.

Table A.2.5 Incidence and Statistical Tests of Carcinogenicity in Male Mice

	Incidence						Significance Levels				
	Veh 1	Veh 2	Low	Med	Med-Hi	Hi	Trend	High vsVeh	MedHi vsVeh	Med vsVeh	Low vsVeh
Male Mice											
ADRENAL, CORTEX											
B-CORTICAL ADENOMA	0	0	1	1	0	1	0.204	0.290	.	0.318	0.328
B-SUBCAPSULAR CELL ADENOMA	4	3	5	3	0	0	0.998	0.935	0.935	0.942	1.000
Subcaps+cortical adenoma	4	3	6	4	0	1	0.955	0.729	0.935	0.477	0.212
ADRENAL, MEDULLA											
B-PHEOCHROMOCYTOMA	1	0	0	0	1	0	0.407	0.290	0.536	0.318	0.328
BRAIN											
M-MENINGEAL SARCOMA	0	0	0	0	0	1	0.148	0.296	.	.	.
CAVITY, ABDOM											
M-MALIGNANT SCHWANNOMA	1	0	0	0	0	0	0.647	0.290	0.318	0.318	0.328
CAVITY, NASAL											
B-ODONTOMA	0	0	0	0	1	0	0.309	.	0.318	.	.

Table A.2.5 (cont.) Incidence and Statistical Tests of Carcinogenicity in Male Mice

	Incidence						Significance Levels				
	Veh 1	Veh 2	Low	Med	Med-Hi	High	Trend	High vsVeh	MedHi vsVeh	Med vsVeh	Low vsVeh
CAVITY, THORACIC											
B-AORTIC BODY TUMOR	0	1	0	0	0	0	0.647	0.290	0.318	0.318	0.328
CECUM											
M-HEMANGIOSARCOMA	0	1	0	0	0	0	0.647	0.290	0.318	0.318	0.328
DUODENUM											
B-ADENOMA	0	0	0	1	0	0	0.309	.	.	0.318	.
FOOT/FOOTPAD											
M-HEMANGIOSARCOMA	0	0	0	1	0	0	0.308	.	.	0.323	.
HARDERIAN GLAND											
B-ADENOMA	9	6	8	4	4	8	0.249	0.385	0.777	0.792	0.503
HEMATO NEOPLASIA											
M-HISTIOCYTIC SARCOMA	1	2	0	0	0	0	0.956	0.643	0.683	0.683	0.697
M-LYMPHOMA, MALIGNANT	2	3	1	2	4	4	0.096	0.270	0.309	0.381	0.634
KIDNEY											
Adenoma/Carc. Tubular	0	1	1	2	0	0	0.789	0.290	0.318	0.237	0.550
B-TUBULAR CELL ADENOMA	0	0	0	2	0	0	0.524	.	.	0.099	.
M-HEMANGIOSARCOMA	0	0	0	0	1	0	0.309	.	0.318	.	.
M-TUBULAR CELL CARCINOMA	0	1	1	0	0	0	0.754	0.290	0.318	0.318	0.550
LIVER											
Adenoma/Carc. Hepato.	10	9	7	3	2	3	0.958	0.942	0.987	0.963	0.691
B-HEPATOCELL ADENOMA, MULT	0	0	1	1	0	0	0.550	.	.	0.318	0.328
B-HEPATOCELLULAR ADENOMA	3	3	2	0	0	0	0.998	0.879	0.905	0.905	0.522
M-HEMANGIOSARCOMA	1	3	0	2	0	1	0.613	0.465	0.788	0.633	0.801
M-HEPATOCELLULAR CARCINOMA	8	7	4	2	2	3	0.824	0.841	0.952	0.952	0.826
LUNG											
B-BRONCHIOLAR/ALVEOLAR ADENO	8	11	5	8	4	5	0.774	0.760	0.917	0.548	0.857
B-BRONCHIOLAR/ALVEOLAR ADENO	0	3	1	1	0	0	0.924	0.646	0.686	0.380	0.398
Bronch/alv. Adenoma/Carc.	13	20	8	10	6	6	0.964	0.975	0.991	0.912	0.969
M-BRONCH/ALVEO CARCIN, MULT	1	0	0	0	0	0	0.647	0.290	0.318	0.318	0.328
M-BRONCHIO/ALVEOLAR CARCINOM	5	6	2	1	2	1	0.923	0.912	0.843	0.940	0.862
MUSCLE, SKELETAL											
M-FIBROSARCOMA	0	0	0	0	0	1	0.148	0.296	.	.	.
M-HEMANGIOSARCOMA	0	0	1	0	0	0	0.474	.	.	.	0.328
NERVE, OPTIC											
M-MALIGNANT SCHWANNOMA	0	0	1	0	0	0	0.474	.	.	.	0.328
PINNA											
M-NEUROFIBROSARCOMA	1	0	0	0	0	0	0.647	0.290	0.318	0.318	0.328
PITUITARY											
Adenoma pars dista/inter	1	0	0	1	1	1	0.200	0.506	0.536	0.536	0.328
B-ADENOMA, PARS DISTALIS	1	0	0	0	0	1	0.275	0.506	0.318	0.318	0.328
B-ADENOMA, PARS INTERMEDIA	0	0	0	1	1	0	0.295	.	0.318	0.318	.
PROSTATE											
B-ADENOMA	0	1	0	0	0	0	0.647	0.290	0.318	0.318	0.328
SEMINAL VESICLE											
M-ADENOCARCINOMA	0	1	0	0	0	0	0.647	0.290	0.318	0.318	0.328
M-SARCOMA, UNDIFFERENTIATED	0	0	1	0	1	0	0.350	.	0.318	.	0.328
SKIN											
B-PAPILLOMA, SQUAMOUS CELL	0	0	0	1	0	0	0.309	.	.	0.318	.
SPLEEN											
M-HEMANGIOSARCOMA	0	0	0	0	0	1	0.148	0.296	.	.	.
SUBCUTANEOUS TIS											
M-FIBROSARCOMA	0	1	0	1	1	2	0.059	0.209	0.536	0.536	0.328
M-HEMANGIOSARCOMA	0	0	0	0	0	1	0.148	0.296	.	.	.
M-MALIGNANT SCHWANNOMA	0	1	0	0	0	0	0.647	0.290	0.318	0.318	0.328
Systemic											
B-HEMANGIOMA	0	1	0	0	1	0	0.407	0.290	0.536	0.318	0.328
Hemangioma/-sarcoma	1	5	2	3	2	3	0.336	0.524	0.489	0.610	0.516
M-HEMANGIOSARCOMA	1	4	2	3	1	3	0.294	0.439	0.622	0.522	0.417

Table A.2.5 (cont.) Incidence and Statistical Tests of Carcinogenicity in Male Mice

	Incidence						Significance Levels				
	Veh 1	Veh 2	Low	Med	Med-Hi	High	Trend	High vsVeh	MedHi vsVeh	Med vsVeh	Low vsVeh
TAIL											
B-KERATOACANTHOMA	0	0	0	0	1	0	0.309	.	0.318	.	.
TESTIS											
B-HEMANGIOMA	0	1	0	0	1	0	0.407	0.290	0.536	0.318	0.328
B-INTERSTITIAL CELL TUMOR	1	2	1	0	0	0	0.955	0.646	0.686	0.686	0.398
M-MALIGNANT SCHWANNOMA	1	0	1	0	0	0	0.754	0.290	0.318	0.318	0.550
THYROID											
B-FOLLICULAR CELL ADENOMA	1	0	1	1	0	0	0.733	0.290	0.318	0.536	0.550
M-FOLLICULAR CELL CARCINOMA	0	0	0	1	0	0	0.309	.	.	0.318	.
TONGUE											
M-SQUAMOUS CELL CARCINOMA	1	0	0	0	0	0	0.647	0.290	0.318	0.318	0.328
WHOLE BODY											
M-HEMANGIOSARCOMA	0	0	1	0	0	0	0.474	.	.	.	0.328

Table A.2.6 Incidence and Statistical Tests of Carcinogenicity in Female Mice

	Incidence						Significance Levels				
	Veh 1	Veh 2	Low	Med	Med-Hi	High	Trend	High vsVeh	MedHi vsVeh	Med vsVeh	Low vsVeh
ADRENAL, CORTEX											
B-SUBCAPSULAR CELL ADENOMA	0	0	0	1	0	1	0.117	0.286	.	0.333	.
M-SUBCAPSULAR CELL CARCINOMA	0	0	1	0	0	0	0.476	.	.	.	0.313
Suncaps Adenoma/Carcinoma	0	0	1	1	0	1	0.198	0.286	.	0.333	0.313
ADRENAL, MEDULLA											
B-PHEOCHROMOCYTOMA	0	0	0	0	3	0	0.378	.	0.026	.	.
CERVIX											
B-ENDOMETRIAL STROMAL POLYP	0	1	0	0	0	0	0.639	0.286	0.297	0.333	0.308
B-LEIOMYOMA	1	0	1	0	0	0	0.754	0.286	0.297	0.333	0.522
M-MALIGNANT SCHWANNOMA	0	1	0	0	0	0	0.639	0.286	0.297	0.333	0.308
FOOT/FOOTPAD											
B-HEMANGIOMA	0	0	0	1	0	0	0.297	.	.	0.333	.
GALLBLADDER											
B-ADENOMA, PAPILLARY	2	0	0	0	0	0	0.870	0.491	0.507	0.557	0.522
HARDERIAN GLAND											
Adenoma/Carcinoma	5	3	7	3	5	8	0.035	0.049	0.341	0.538	0.138
B-ADENOMA	5	3	7	2	5	7	0.067	0.098	0.341	0.704	0.138
M-CARCINOMA	0	0	0	2	0	1	0.174	0.286	.	0.110	.
HEMATO NEOPLASIA											
M-HISTIOCYTIC SARCOMA	3	6	1	0	3	0	0.945	0.953	0.497	0.975	0.870
M-LYMPHOMA, MALIGNANT	11	12	13	6	5	10	0.445	0.542	0.906	0.898	0.347
M-MYELOGENOUS LEUKEMIA	0	0	1	1	0	0	0.537	.	.	0.333	0.313
LIVER											
B-HEPATOCELLULAR ADENOMA	0	1	0	0	0	0	0.639	0.286	0.297	0.333	0.308
M-HEPATOCELLULAR CARCINOMA	0	1	0	1	1	0	0.470	0.286	0.515	0.557	0.308
M-OSTEOSARCOMA	0	0	1	0	0	0	0.476	.	.	.	0.313
LUNG											
B-BRONCH/ALVEO ADENOMA, MULT	0	0	0	0	0	1	0.145	0.286	.	.	.
B-BRONCHIO/ALVEOLAR ADENOMA	4	7	6	5	4	2	0.894	0.781	0.498	0.453	0.451
Bronch/alv. Adenoma/Carc.	8	12	12	8	7	4	0.951	0.875	0.622	0.640	0.261
M-BRONCHI/ALVEOLAR CARCINOMA	4	5	7	3	3	1	0.951	0.830	0.504	0.600	0.195
M-FIBROSARCOMA	0	0	0	1	0	0	0.297	.	.	0.333	.
MAMMARY, FEMALE											
Adenoma/Carc./-sarcoma/-canth.	2	1	13	19	26	20	0.000	0.000	0.000	0.000	0.000
B-ADENOMA	0	0	0	0	1	0	0.297	.	0.297	.	.
B-ADENOMA, PAPILLARY DUCT	0	0	0	0	0	1	0.145	0.286	.	.	.
M-ADENOACANTHOMA	1	0	7	6	7	5	0.103	0.008	0.001	0.006	0.001
M-CARCINOMA	2	1	7	12	18	13	0.001	0.000	0.000	0.000	0.011
M-CARCINOSARCOMA	0	0	0	1	2	2	0.026	0.080	0.090	0.333	.
M-HEMANGIOSARCOMA	1	0	0	0	0	0	0.639	0.286	0.297	0.333	0.308

Table A.2.6 (cont.) Incidence and Statistical Tests of Carcinogenicity in Female Mice

	Incidence						Significance Levels				
	Veh 1	Veh 2	Low	Med	Med-Hi	Hi	Trend	High vsVeh	MedHi vsVeh	Med vsVeh	Low vsVeh
MUSCLE, SKELETAL											
M-FIBROSARCOMA	0	0	0	0	0	1	0.145	0.286	.	.	.
M-RHABDOMYOSARCOMA	0	2	0	0	0	0	0.870	0.491	0.507	0.557	0.522
NERVE, SCIATIC											
M-MALIGNANT SCHWANNOMA	1	0	0	0	0	0	0.639	0.286	0.297	0.333	0.308
OVARY											
B-CYSTADENOMA	1	0	1	2	1	1	0.325	0.491	0.507	0.258	0.522
B-HEMANGIOMA	0	0	0	1	0	0	0.297	.	.	0.333	.
B-LUTEOMA	1	1	0	0	0	0	0.870	0.491	0.507	0.557	0.522
B-THECAL CELL TUMOR	0	0	0	1	0	0	0.297	.	.	0.333	.
B-TUBULOSTROMAL ADENOMA	0	0	1	0	0	0	0.476	.	.	.	0.313
Cystad./Gran./Thecal/Tubul.	1	0	2	4	1	2	0.254	0.196	0.507	0.045	0.231
M-GRANULOSA CELL TUMOR, MALI	0	0	0	1	0	1	0.117	0.286	.	0.333	.
PANCREAS											
B-ACINAR CELL ADENOMA	0	0	0	1	0	0	0.297	.	.	0.333	.
B-ISLET CELL ADENOMA	0	0	0	1	3	1	0.155	0.286	0.026	0.333	.
PARATHYROID											
B-ADENOMA	1	0	0	0	0	0	0.639	0.286	0.297	0.333	0.308
PITUITARY											
B-ADENOMA, PARS DISTALIS	3	4	11	17	27	29	0.000	0.000	0.000	0.000	0.007
SKIN											
B-KERATOACANTHOMA	0	1	1	0	2	0	0.509	0.284	0.207	0.331	0.519
B-MASTOCYTOMA	0	0	0	0	1	0	0.297	.	0.297	.	.
B-PAPILLOMA, SQUAMOUS CELL	1	0	0	1	0	0	0.597	0.286	0.297	0.557	0.308
B-SEBACEOUS CELL ADENOMA	0	0	0	0	0	1	0.145	0.286	.	.	.
M-BASOSQUAMOUS CELL CARCINOM	0	0	0	0	1	0	0.297	.	0.297	.	.
M-FIBROSARCOMA	0	0	0	0	0	1	0.145	0.286	.	.	.
M-SQUAMOUS CELL CARCINOMA	0	0	0	1	0	0	0.297	.	.	0.333	.
sq. cell/basosq./kerato.	1	1	1	2	3	0	0.733	0.488	0.152	0.403	0.668
SPLEEN											
B-HEMANGIOMA	0	0	0	1	0	0	0.297	.	.	0.333	.
M-HEMANGIOSARCOMA	0	1	0	0	0	0	0.639	0.286	0.297	0.333	0.308
SUBCUTANEOUS TIS											
M-FIBROSARCOMA	0	0	1	0	0	0	0.476	.	.	.	0.313
M-OSTEOSARCOMA	0	1	0	0	0	0	0.636	0.284	0.295	0.331	0.305
M-RHABDOMYOSARCOMA	0	1	0	0	0	0	0.639	0.286	0.297	0.333	0.308
M-SARCOMA, UNDIFFERENTIATED	0	0	0	0	0	1	0.145	0.286	.	.	.
Systemic											
B-HEMANGIOMA	0	1	0	3	0	0	0.735	0.286	0.297	0.108	0.308
Hemangioma/-sarcoma	1	3	2	3	0	1	0.745	0.439	0.757	0.424	0.607
M-HEMANGIOSARCOMA	1	2	2	0	0	1	0.583	0.318	0.652	0.704	0.495
TAIL											
M-HEMANGIOSARCOMA	0	0	0	0	0	1	0.145	0.286	.	.	.
THYROID											
B-FOLLICULAR CELL ADENOMA	0	0	1	0	0	0	0.476	.	.	.	0.313
UTERUS											
B-ENDOMETRIAL STROMAL POLYP	2	2	1	0	1	0	0.871	0.741	0.462	0.804	0.485
B-HEMANGIOMA	0	1	0	0	0	0	0.639	0.286	0.297	0.333	0.308
B-LEIOMYOMA	1	1	1	0	1	0	0.692	0.491	0.656	0.557	0.672
M-ADENOCARCINOMA	1	0	0	0	0	0	0.639	0.286	0.297	0.333	0.308
M-HEMANGIOSARCOMA	0	0	1	0	0	0	0.476	.	.	.	0.313
M-LEIOMYOSARCOMA	1	1	1	0	0	0	0.891	0.491	0.507	0.557	0.679
M-MALIGNANT SCHWANNOMA	1	1	0	0	0	0	0.870	0.491	0.507	0.557	0.522
M-SARCOMA, UNDIFFERENTIATED	1	0	0	0	0	0	0.639	0.286	0.297	0.333	0.308
VAGINA											
B-SQUAMOUS CELL PAPILLOMA	0	0	0	0	0	1	0.145	0.286	.	.	.
WHOLE BODY											
M-HEMANGIOSARCOMA	0	1	1	0	0	0	0.751	0.284	0.295	0.331	0.526

Appendix 3. References

Bailer, A. and Portier, C. (1988), "Effects of Treatment-Induced Mortality on Tests for Carcinogenicity in Small Samples", *Biometrics*, **44**, 4, 417-431.

Bieler, G.S., and Williams, R.L. (1993), "Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity", *Biometrics*, **49**, 4, 793-801.

Chu, K.C., Ceuto, C., and Ward, J.M. (1981), Factors in the Evaluation of 200 National Cancer Institute Carcinogen Bioassays, *Journal of Toxicology and Environmental Health,* 8, 251-280.

Haseman, J. K. (1983), A Reexamination of False-positive Rates for Carcinogenicity Studies, *Fundamental and Applied Toxicology,* 3, 334-339.

Lin, K. K. and Ali, M.W. (2006), Statistical Review and Evaluation of Animal Tumorigenicity Studies, *Statistics in the Pharmaceutical Industry,Third Edition,* edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.

Lin, K. K. and Rahman, M.A. (1998), Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs, *Journal of Biopharmaceutical Statistics*. 8(1), 1-15.

McConnell, E.E., Solleveld, H.A., Swenberg, J.A., and Boorman, G.A. (1986), Guidelines for Combining Neoplasms for Evaluation of Rodent Carcinogenesis Studies, *Journal of the National Cancer Institute*. 76, 283-289.

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments, *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal*, International Agency for Research Against Cancer, 311-426.

Rahman, M.A. and Lin, K.K. (2008), A Comparison of False Positive Rates of Peto and Poly-3 Methods for Long Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman, *Journal of Biopharmaceutical Statistics*. 18, 949-958.

STP Peto Working Group (2002), Statistical Methods for Carcinogenicity Studies, *Toxicologic Pathology*. 30 (3), 403-414.

U.S. Department of Health and Human Services (2001), Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies

of Pharmaceuticals (DRAFT GUIDANCE), Center for Drug Evaluation and Research, Food and Drug Administration.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-200603	ORIG-1	DAINIPPON SUMITOMO PHARMA AMERICA INC	Lurasidone HCl

---

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

/s/

STEVEN F THOMSON
06/28/2010

KARL K LIN
06/29/2010
Concur with review

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 200603 **Applicant: Dainippon Sumitomo** **Stamp Date: 12/30/2009**

Drug Name: Lurasidone **NDA/BLA Type: O-1**

On **initial** overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	X			Except efficacy subgroup analysis for studies D1050006 and D1050196
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? ___Yes___

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			X	
Appropriate references for novel statistical methodology (if present) are included.	X			
Safety data organized to permit analyses across clinical trials in the NDA/BLA.			X	
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			Response profile plots are requested.

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

George Kordzakhia	March 8, 2010
Reviewing Statistician	Date
Peiling Yang	
Supervisor/Team Leader	Date

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-200603	ORIG-1	DAINIPPON SUMITOMO PHARMA AMERICA INC	Lurasidone HCl

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

/s/

GEORGE KORDZAKHIA
03/08/2010

PEILING YANG
03/08/2010