

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

204370Orig1Orig2s000

STATISTICAL REVIEW(S)

ADDENDUM

CLINICAL STUDIES

NDA #: 204-370 (Original)
Drug Name: Cariprazine
Indications: Schizophrenia and Bipolar I Disorder
Applicant: Forest
Dates: Stamp date: 11/19/2012; Filing date: 1/18/2013; PDUFA due date 11/19/2013
Review Priority: Standard

Biometrics Division: Division of Biometrics I
Statistical Reviewer: Eiji Ishida, M.S.
Concurring Reviewers: Peiling Yang, Ph.D.; H. M. James Hung, Ph.D.

Medical Division: Division of Psychiatry Products
Clinical Team: Francis Becker, M.D., Robert Levin, M.D.
Project Manager: Kimberly Updegraff, Pharm.D.

Keywords: NDA review, Multiple endpoints, Multiple doses

This Addendum corrects a few typos appearing in my review signed off on July 22, 2013. The corrections are highlighted in red. They do not affect the review and its conclusion.

(1) Section 3.2.4.1.2 Reviewer’s assessments

Table 8 of the original review (page 22), as appears below:

Table 8: MMRM and ANCOVA (LOCF) results for primary efficacy (PANSS total score)

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-04 (N=454)	Cariprazine 3 mg/day (N=151)	96.1 (8.7)	-20.2 (1.5)	-6.0	-16.4 (1.4)	-5.4
	Cariprazine 6 mg/day (N=154)	95.7 (9.4)	-23.0 (1.5)	-8.8	-18.9 (1.4)	-7.9
	Aripirazole 10 mg/day (N=150)	95.6 (9.0)	-21.2 (1.4)	-7.0	-18.8 (1.4)	-7.7
	Placebo (N=149)	96.5 (9.1)	-14.3 (1.5)	-	-11.0 (1.4)	-
MD-05 (N=439)	Cariprazine 3-6 mg/day (N=147)	96.3 (9.3)	-22.8 (1.6)	-6.8	-18.6 (1.5)	-6.3
	Cariprazine 6-9 mg/day (N=147)	96.3 (9.0)	-25.9 (1.7)	-9.9	-20.2 (1.5)	-8.0
	Placebo (N=145)	96.2 (9.3)	-16.0 (1.6)	-	-12.3 (1.5)	-
MD-16 (N=573)	Cariprazine 1.5 mg/day (N=140)	97.1 (9.1)	-21.3 (1.8)	-8.0	-19.4 (1.6)	-7.6
	Cariprazine 3 mg/day (N=140)	97.2 (8.7)	-21.5 (1.7)	-8.2	-20.7 (1.6)	-8.9
	Cariprazine 4.5 mg/day (N=145)	96.7 (9.0)	-23.8 (1.7)	-10.5	-22.3 (1.6)	-10.5
	Risperidone 4 mg/day (N=138)	98.1 (9.4)	-29.3 (1.7)	-16.0	-27.0 (1.6)	-15.2
	Placebo (N=148)	97.3 (9.2)	-13.3 (1.8)	-	-11.8 (1.5)	-
MD-03 (N=377)	Cariprazine 1.5-4.5 mg/day (N=122)					
	Cariprazine 6-12 mg/day (N=129)					
	Placebo (N=126)					

[Source: Sponsor’s Clinical Study Reports and Reviewer’s results]

Note: (1) The bolded results are from pre-specified primary analysis. (2) Individual unadjusted 95% confidence intervals for mean difference from placebo are provided in Table 9 and Figure 1. (3) Adjusted *p* values for the primary endpoints (Studies MD-04, MD-05 and MD-16) are given in Table 6.

should be replaced by the following corrected table:

Table 8: MMRM and ANCOVA (LOCF) results for primary efficacy (PANSS total score)

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-04 (N=454)	Cariprazine 3 mg/day (N=151)	96.1 (8.7)	-20.2 (1.5)	-6.0	-16.4 (1.4)	-5.4
	Cariprazine 6 mg/day (N=154)	95.7 (9.4)	-23.0 (1.5)	-8.8	-18.9 (1.4)	-7.9
	Aripirazole 10 mg/day (N=150)	95.6 (9.0)	-21.2 (1.4)	-7.0	-18.8 (1.4)	-7.7
	Placebo (N=149)	96.5 (9.1)	-14.3 (1.5)	-	-11.0 (1.4)	-
MD-05 (N=439)	Cariprazine 3-6 mg/day (N=147)	96.3 (9.3)	-22.8 (1.6)	-6.8	-18.6 (1.5)	-6.3
	Cariprazine 6-9 mg/day (N=147)	96.3 (9.0)	-25.9 (1.7)	-9.9	-20.2 (1.5)	-8.0
	Placebo (N=145)	96.6 (9.3)	-16.0 (1.6)	-	-12.3 (1.5)	-
MD-16 (N=573)	Cariprazine 1.5 mg/day (N=140)	97.1 (9.1)	-21.3 (1.8)	-8.0	-19.4 (1.6)	-7.6
	Cariprazine 3 mg/day (N=140)	97.2 (8.7)	-21.5 (1.7)	-8.2	-20.7 (1.6)	-8.9
	Cariprazine 4.5 mg/day (N=145)	96.7 (9.0)	-23.8 (1.7)	-10.5	-22.3 (1.6)	-10.5
	Risperidone 4 mg/day (N=138)	98.1 (9.4)	-29.3 (1.7)	-16.0	-27.0 (1.6)	-15.2
	Placebo (N=148)	97.3 (9.2)	-13.3 (1.8)	-	-11.8 (1.5)	-
MD-03 (N=377)	Cariprazine 1.5-4.5 mg/day (N=122)					
	Cariprazine 6-12 mg/day (N=129)					
	Placebo (N=126)					

(b) (4)

[Source: Sponsor's Clinical Study Reports and Reviewer's results]

Note: (1) The bolded results are from pre-specified primary analysis. (2) Individual unadjusted 95% confidence intervals for mean difference from placebo are provided in Table 9 and Figure 1. (3) Adjusted *p* values for the primary endpoints (Studies MD-04, MD-05 and MD-16) are given in Table 6.

(2) Section 3.2.4.2.2 Reviewer's assessments

Table 13 of the original review (page 32), as appears below:

Table 13: MMRM and ANCOVA (LOCF) results for primary efficacy (YMRS total score)

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-31 (N=235)	Cariprazine 3-12 mg/day (N=118)	30.6 (5.0)	-15.5 (1.1)	-7.0	-13.3 (1.1)	-6.1
	Placebo (N=117)	30.2 (5.2)	-8.5 (1.1)	-	-7.2 (1.1)	-
MD-32 (N=310)	Cariprazine 3-12 mg/day (N=158)	32.3 (5.8)	-19.6 (0.9)	-4.3	-17.3 (0.9)	-4.3
	Placebo (N=152)	32.1 (5.6)	-15.3 (0.9)	-	-13.0 (0.9)	-
MD-33 (N=492)	Cariprazine 3-6 mg/day (N=165)	33.2 (5.6)	-18.6 (0.8)	-6.1	-17.9 (0.8)	-5.8
	Cariprazine 6-12 mg/day (N=167)	33.0 (4.7)	-18.5 (0.8)	-5.9	-17.4 (0.8)	-5.3
	Placebo (N=160)	32.6 (5.8)	-12.5 (0.8)	-	-12.1 (0.8)	-

[Source: Sponsor's Clinical Study Reports and Reviewer's results]

Note: (1) The bolded results are from pre-specified primary analysis. (2) In MMRM analysis of Study RGH-MD-31, an interaction term of Baseline score and Visit was not included in the model. This reviewer confirmed the results are almost identical when this interaction term is included. (3) Individual unadjusted 95% confidence intervals for mean difference from placebo are provided in Table 14 and Figure 3. (In the MMRM analysis of Study MD-31, this reviewer used a model that includes an interaction term of Baseline score and Visit, and thus the LS mean difference from placebo was different from that of Table 13. The difference was negligible.) (4) Adjusted *p* values for the primary endpoints are given in Table 11.

should be replaced by the following corrected table (displayed in the next page):

Table 13: MMRM and ANCOVA (LOCF) results for primary efficacy (YMRS total score)

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-31 (N=235)	Cariprazine 3-12 mg/day (N=118)	30.6 (5.0)	-15.5 (1.1)	-7.0	-15.0 (1.1)	-6.1
	Placebo (N=117)	30.2 (5.2)	-8.5 (1.1)	-	-8.9 (1.1)	-
MD-32 (N=310)	Cariprazine 3-12 mg/day (N=158)	32.3 (5.8)	-19.6 (0.9)	-4.3	-17.3 (0.9)	-4.3
	Placebo (N=152)	32.1 (5.6)	-15.3 (0.9)	-	-13.0 (0.9)	-
MD-33 (N=492)	Cariprazine 3-6 mg/day (N=165)	33.2 (5.6)	-18.6 (0.8)	-6.1	-17.9 (0.8)	-5.8
	Cariprazine 6-12 mg/day (N=167)	33.0 (4.7)	-18.5 (0.8)	-5.9	-17.4 (0.8)	-5.3
	Placebo (N=160)	32.6 (5.8)	-12.5 (0.8)	-	-12.1 (0.8)	-

[Source: Sponsor's Clinical Study Reports and Reviewer's results]

Note: (1) The bolded results are from pre-specified primary analysis. (2) In MMRM analysis of Study RGH-MD-31, an interaction term of Baseline score and Visit was not included in the model. This reviewer confirmed the results are almost identical when this interaction term is included. (3) Individual unadjusted 95% confidence intervals for mean difference from placebo are provided in Table 14 and Figure 3. (In the MMRM analysis of Study MD-31, this reviewer used a model that includes an interaction term of Baseline score and Visit, and thus the LS mean difference from placebo was different from that of Table 13. The difference was negligible.) (4) Adjusted *p* values for the primary endpoints are given in Table 11.

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

/s/

EIJI ISHIDA
10/08/2013

PEILING YANG
10/08/2013

HSIEN MING J HUNG
10/08/2013



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 #: 204-370 (Original)
Drug Name: Cariprazine
Indications: Schizophrenia and Bipolar I Disorder
Applicant: Forest
Dates: Stamp date: 11/19/2012; Filing date: 1/18/2013; PDUFA due date 11/19/2013
Review Priority: Standard

Biometrics Division: Division of Biometrics I
Statistical Reviewer: Eiji Ishida, M.S.
Concurring Reviewers: Peiling Yang, Ph.D.; H. M. James Hung, Ph.D.

Medical Division: Division of Psychiatry Products
Clinical Team: Francis Becker, M.D., Robert Levin, M.D.
Project Manager: Kimberly Updegraff, Pharm.D.

Keywords: NDA review, Multiple endpoints, Multiple doses

Table of Contents

1 EXECUTIVE SUMMARY	3
1.1 CONCLUSIONS AND RECOMMENDATIONS	3
1.1.1 <i>Schizophrenia</i>	3
1.1.2 <i>Bipolar I Disorder</i>	3
1.2 STATISTICAL ISSUES AND FINDINGS	4
1.2.1 <i>Schizophrenia</i>	4
1.2.2 <i>Bipolar I Disorder</i>	5
2 INTRODUCTION.....	5
2.1 OVERVIEW	5
2.1.1 <i>Regulatory History</i>	6
2.1.2 <i>Schizophrenia Program</i>	8
2.1.3 <i>Bipolar I Disorder Program</i>	8
2.2 DATA SOURCES	9
3 STATISTICAL EVALUATION.....	9
3.1 DATA AND ANALYSIS QUALITY	9
3.2 EVALUATION OF EFFICACY.....	10
3.2.1 <i>Study Design and Endpoints</i>	10
3.2.2 <i>Statistical Methodologies</i>	11
3.2.3 <i>Patient Disposition, Demographic and Baseline Characteristics</i>	13
3.2.4 <i>Primary Analysis - Results and Conclusions</i>	16
3.3 EVALUATION OF SAFETY	41
4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	41
4.1 GENDER, RACE, AGE GROUP, AND COUNTRY	41
4.1.1 <i>Schizophrenia Pivotal Studies</i>	41
4.1.2 <i>Bipolar I Disorder Pivotal Studies</i>	48
5 SUMMARY AND CONCLUSIONS.....	53
5.1 STATISTICAL ISSUES	53
5.1.1 <i>Schizophrenia</i>	53
5.1.2 <i>Bipolar I Disorder</i>	54
5.2 CONCLUSIONS AND RECOMMENDATIONS	55
5.2.1 <i>Schizophrenia</i>	55
5.2.2 <i>Bipolar I Disorder</i>	55

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

1.1.1 Schizophrenia

The cariprazine schizophrenia NDA program of Forest includes three pivotal studies and one supportive study for efficacy assessment of the candidate drug. Two pivotal fixed dose studies of this program have statistically established evidence that daily fixed doses of cariprazine 1.5 mg, 3 mg, 4.5 mg and 6 mg are efficacious. One pivotal fixed-flexible dose study of the program has established statistical evidence that two cariprazine dose ranges of 3-6 mg and 6-9 mg were efficacious in optimized dose within the fixed dose range. Another proof-of-concept, fixed-flexible dose study with two dose-ranges 1.5-4.5 mg and 6-12 mg, submitted as a supportive study, has shown a trend for cariprazine efficacy, although it failed in showing statistical significance.

The sponsor considers doses of 1.5 to 9 mg administered once daily effective. (b) (4)

Based on statistical evidence, this reviewer suggests that the doses of 1.5 mg, 3 mg, 4.5 mg and 6 mg may be included in the product label as *effective doses*, as an effective dose is usually meant for efficacy of a fixed daily dose in the schizophrenia product label. A dose response relationship has not been consistently observed across the two fixed dose studies.

Given the positive results of the fixed-flexible dose study, an approval of the two dose-ranges as *effective dose ranges* may also be considered. The appropriateness of such a decision may be a clinical matter. It may not be appropriate to draw any conclusion on a dose response relationship, based on the efficacy estimates of the dose ranges (3-6 mg and 6-9 mg).

1.1.2 Bipolar I Disorder

The cariprazine bipolar I disorder NDA program of Forest includes three pivotal studies, i.e., two flexible dose studies and one fixed-flexible dose study for efficacy assessment of the candidate drug. A dose-range of cariprazine 3-12 mg has been statistically shown efficacious in the two pivotal flexible dose studies. Furthermore, two dose-ranges of cariprazine 3-6 mg and 6-12 mg have been statistically shown efficacious in the pivotal fixed-flexible dose study.

It may be a clinical matter how efficacy of the dose-ranges should be described in the label. As no efficacy data of fixed doses in bipolar mania are available, the submitted data do not admit a statistical assessment of efficacy of fixed doses.

This reviewer cautions that it may not be appropriate to draw any conclusion on a dose response relationship based on efficacy results of the dose ranges. An efficacy comparison of dose ranges may not be meaningful when individual doses were optimized. At any rate, the efficacy estimates of the three dose ranges (3-6 mg, 6-12 mg and 3-12 mg) did not show a consistent dose response relationship.

1.2 Statistical Issues and Findings

1.2.1 Schizophrenia

In the schizophrenia program, three pivotal studies have shown efficacy of cariprazine:

- In two fixed dose studies, efficacy of daily fixed doses of cariprazine 1.5 mg, 3 mg, 4.5 mg and 6 mg was established at the 6-week endpoint based on PANSS total score and CGI-S score.
- In one fixed-flexible dose study, efficacy of cariprazine dose ranges of 3-6 mg and 6-9 mg was established at the 6-week endpoint based on PANSS total score and CGI-S score.

One supportive study did not achieve statistical significance. However, it has shown a trend for efficacy of cariprazine of a dose range of 3-12 mg at the 6-week endpoint based on PANSS total score and CGI-S score.

It is noted that efficacy of a daily fixed dose of 9 mg has not been fully evaluated in the clinical program, although there were some patients in the phase-3 fixed-flexible dose study who reached the upper-end dose (9 mg) of a dose range of 6-9 mg for their optimized dose. The efficacy estimates from a fixed-flexible dose study may not serve as a basis for establishing efficacy of fixed doses. In dose optimization as conducted in the fixed-flexible dose studies, it may not be feasible to evaluate “efficacy difference due to different doses”. By design, only part of randomized patients assigned to the 6-9 mg dose range reached 9 mg for their optimal dose (Table 7). In general, observed efficacy of an optimal dose may not be sufficient for understanding efficacy of a fixed dose. Furthermore, the dose range was not designed to be compared with any fixed dose within the same study. This reviewer notes that the FDA communicated to the sponsor its preference for studies looking at fixed doses within the proposed fixed-flexible dose study, at Type C Guidance meeting held on February 11, 2009¹. This reviewer also points out that, at this meeting, (b) (4)

(b) (4)
This reviewer notes that the proportion of subjects who reached the upper-end dose (of a dose-range), their efficacy outcome and overall mean daily exposure were not consistent between the two fixed-flexible dose studies (Table 7). See *Section 3.2.4.1.2 Reviewer’s assessments* for more details.

The efficacy estimates of fixed doses of 1.5 mg, 3 mg, 4.5 mg and 6 mg do not seem to suggest a dose-response relationship across the two fixed dose studies. Dose response may have been observed in each of the two fixed dose studies, but no consistent dose responses for the fixed doses were observed across these studies (Figure 1 and Figure 2).

¹ Preliminary Comments/Minutes of Meeting IND (b) (4) [Schizophrenia], IND 77726 [bipolar]: RGH-188 (cariprazine). The meeting minutes document is available in DARRTS. It was finalized on February 20, 2009.

In the fixed-flexible dose study (MD-05), the efficacy estimate of 6-9 mg was better than that of 3-6 mg (Table 8). However, an efficacy comparison of dose ranges may not be meaningful when dose was optimized.

1.2.2 Bipolar I Disorder

This reviewer notes that no information on efficacy of fixed doses is available in the submitted NDA study data. At Type C meeting (see footnote 1), the sponsor asked whether it would be sufficient to demonstrate efficacy in two flexible dose studies in mania, as opposed to one fixed-dose study and one flexible-dose study. The FDA cautioned that assessing effective doses without fixed dose studies may become difficult, but agreed that it would be acceptable for the sponsor to conduct one flexible dose study in mania and one fixed-flexible dose study in mania.

A dose response relationship of fixed doses may be difficult to assess from the submitted efficacy data, since there is no information on efficacy of fixed doses in the bipolar studies. See Section 3.2.4.2.2 *Reviewer's assessments* for more details.

A differential in efficacy estimates between the two dose-ranges, 3-6 mg and 6-12 mg, may be viewed as negligible. The efficacy estimate of the 3-12 mg dose range was much better in MD-32 than in MD-31, but the proportion of subjects of 3-12 mg dose-range who reached the target (upper-end) dose of 12 mg was much higher in MD-31 than in MD-32 (See Table 12 and Table 13). Furthermore, dose increase decisions were made differently in the same dose range (3-12 mg) between the two studies MD-31 and MD-32 (See Section 3.2.4.2.2 *Reviewer's assessments*). The two dose-ranges of Study MD-33 (3-6 mg and 6-12 mg) exhibited almost no difference in efficacy estimates. This may be natural as dose was optimized. It may not be meaningful to draw a conclusion on dose response of the dose-ranges.

Note: In both indications, the above efficacy conclusions are based on statistical evidence obtained from efficacy studies whose study designs and analysis plans were agreed upon between the sponsor and the FDA. The efficacy estimates of the primary and key secondary endpoints are displayed in Table 8 and Table 10 for schizophrenia and in Table 13 and Table 15 for bipolar mania.

2 INTRODUCTION

2.1 Overview

The sponsor submitted three pivotal studies and one supportive study for efficacy evidence to support their NDA for cariprazine for the treatment of schizophrenia, and three pivotal studies for the treatment of manic and mixed episodes associated with bipolar I disorder.

In the first sub-section, we briefly describe the regulatory history of the phase-3 programs of the sponsor's clinical development. In the following sub-sections, we provide the brief summaries of the submitted 6 pivotal studies and one supportive study. In Table 1, the core features of designs of the submitted studies, their treatment arms, and the numbers of randomized subjects and ITT subjects are given.

2.1.1 Regulatory History

At Type B Pre-NDA meeting held on May 24, 2012, the FDA concurred on the sponsor's inquiries:

- Does the Division concur that the positive results from the 3 pivotal studies (Studies RGH-MD-04, RGH-MD-05 and RGH-MD-16) and the supportive study (RGH-MD-03) can support the NDA for cariprazine at doses 1.5 to (b)(4) mg/day for the treatment of patients with schizophrenia?
- Does the Division concur that the positive results from three pivotal studies (Studies RGH-MD-31, RGH-MD-32 and RGH-MD-33) can support the NDA for cariprazine at doses 3 to (b)(4) mg/day for the treatment of manic and mixed episodes associated with bipolar I disorder?

At Type C Guidance meeting held on February 11, 2009, the sponsor discussed with the FDA their phase-3 study clinical development plan, based on (1) one completed fixed-flexible study (RGH-MD-03) and an ongoing fixed-dose study (RGH-MD-16) for schizophrenia program, and (2) one completed flexible-dose study (RGH-MD-31) for bipolar mania program². The FDA advised that at least 2 positive, adequate, and well-controlled studies would be needed for the sponsor to file an NDA for an indication. In each of the indications, the sponsor conducted two phase-3 studies. In schizophrenia phase-3 study program, the sponsor conducted a fixed-flexible dose study (RGH-MD-05), and a fixed dose study (RGH-MD-04). The bipolar mania phase-3 study program, consisted of one fixed-flexible dose study (RGH-MD-33), and one flexible dose study (RGH-MD-32).

From statistical viewpoints, the following agreements between the sponsor and the agency may be important to mention here:

Assay sensitivity: In Study RGH-MD-16, the sponsor had used risperidone 4 mg/day as an active comparator. The FDA encouraged using active comparators in all phase-3 trials in order to establish assay sensitivity. And the sponsor used aripiprazole 10 mg/day for an active comparator in Study RGH-MD-04.

Titration: The FDA asked the sponsor if cariprazine requires titration. The sponsor responded that in a cariprazine study, titration is necessary in order to improve tolerability and reduce the risk of adverse events. In all studies, cariprazine subjects started with a daily dose of 1.5 mg. Titration was planned so subjects would reach a certain dose within a few days of their initial dosing: the target fixed doses in fixed studies (RGH-MD-04), and the low-end doses of a given dose range in fixed-flexible dose studies (RGH-MD-05 and RGH-MD-32) and in a flexible dose study (RGH-MD-33). See Table 1.

Rationale for fixed-flexible dosing: The FDA requested clarification on the rationale for the fixed-flexible dosing regimen for the proposed studies (RGH-MD-05 and RGH-MD-33). *The sponsor clarified that the fixed-flexible dose range provides an option to evaluate a wider dose range using fewer dose groups. The dose range within each group offers flexibility to clinicians within a fixed range: increasing dose if response is not adequate and decreasing dose following an increase if there are tolerability issues³.*

² Preliminary Comments/Minutes of Meeting IND (b)(4) [Schizophrenia], IND 77726 [bipolar]: RGH-188 (cariprazine). The meeting minutes document is available in DARRTS. It was finalized on February 20, 2009.

³ Section 5.1.2 of Integrated Summary of Effectiveness (both for schizophrenia and bipolar I disorder)

Table 1: List of study elements of cariprazine efficacy NDA studies

Applicant defined study number (Total number of randomized subjects, that of ITT subjects)	Study Design	Treatment arms	Number of randomized subjects	Number of ITT subjects
Schizophrenia				
RGH-MD-04 (N=617, N=604)	Randomized, Double-blind, Placebo-and active-controlled, Parallel-group, Fixed-dose (Phase 3, <u>Pivotal</u>)	Cariprazine 3 mg/day	155	151
		Cariprazine 6 mg/day	157	154
		Aripiprazole 10 mg/day	152	150
		Placebo	153	149
RGH-MD-05 (N=446, N=439)	Randomized, Double-blind, Placebo-controlled, Parallel-group, Fixed/flexible-dose (Phase 3, <u>Pivotal</u>)	Cariprazine 3-6 mg/day	151	147
		Cariprazine 6-9 mg/day	148	147
		Placebo	147	145
RGH-MD-16 (N=732, N=711)	Randomized, Double-blind, Placebo-and active-controlled, Parallel-group, Fixed-dose (Phase 2b, <u>Pivotal</u>)	Cariprazine 1.5 mg/day	145	140
		Cariprazine 3 mg/day	147	140
		Cariprazine 4.5 mg/day	148	145
		Risperidone 4 mg/day	141	138
		Placebo	151	148
RGH-MD-03 (N=392, N=377)	Randomized, Double-blind, Placebo-controlled, Parallel-group, Flexible-dose (Phase 2, Supportive)	Cariprazine 1.5-4.5 mg/day	128	122
		Cariprazine 6-12 mg/day	134	129
		Placebo	130	126
Bipolar I Disorder				
RGH-MD-31 (N=238, N=235)	Randomized, Double-blind, Placebo-controlled, Parallel-group, Flexible-dose (Phase 2, <u>Pivotal</u>)	Cariprazine 3-12 mg/day	118	118
		Placebo	120	117
RGH-MD-32 (N=312, N=310)	Randomized, Double-blind, Placebo-controlled, Parallel-group, Flexible-dose (Phase 3, <u>Pivotal</u>)	Cariprazine 3-12 mg/day	158	158
		Placebo	154	152
RGH-MD-33 (N=497, N=492)	Randomized, Double-blind, Placebo-controlled, Parallel-group, Fixed/flexible-dose (Phase 3, <u>Pivotal</u>)	Cariprazine 3-6 mg/day	167	165
		Cariprazine 6-12 mg/day	169	167
		Placebo	161	160

Note: Study RGH-MD-03 involved US study centers alone, while in all other studies, foreign study centers were involved.

Primary analysis for the primary and the key secondary efficacy variables: The FDA and the sponsor reached agreements on pre-specifications of the primary analysis method for the proposed phase-3 studies as well as the two phase-2 studies submitted for the NDA filing through discussions at face-to-face meetings and communications. The pre-specifications

included multiplicity adjustment methods, pooling strategy for small centers, the primary analysis method that is valid for the MAR (missing at random) missing data mechanism.

Schizophrenia phase-3 program:

(b) (4)

However, the sponsor did not investigate fixed doses in the fixed-flexible dose study (RGH-MD-05).

Bipolar phase-3 program: According to the sponsor, in the flexible-dose study (RGH-MD-31), the planned dose range of which was 3-12 mg/day, more than 90% of subjects were treated with doses of 6-12 mg/day. The sponsor planned to use a dose range of 6-12 mg/day in a phase-3 flexible dose study. The sponsor asked whether it would be sufficient to demonstrate efficacy in two flexible dose studies in mania, as opposed to one fixed-dose study and one flexible-dose study. The FDA emphasized *the importance of establishing the minimal effective dose; a fixed-dose study would be the ideal study design to meet this objective*. The possibility of a fixed-flexible dose study using fixed ranges of doses was also discussed. The FDA agreed that it would be acceptable for the sponsor to conduct one flexible dose study in mania and one fixed-flexible dose study in mania⁵. It is noted that the sponsor has no fixed dose study in the bipolar program.

2.1.2 Schizophrenia Program

The cariprazine program for the indication of schizophrenia included three pivotal, 6-week, double-blind, placebo-controlled studies in adult patients with acute exacerbation of schizophrenia. Study RGH-MD-04 was a phase-3, randomized, double-blind, placebo- and active-controlled, parallel-group, fixed-dose (two fixed doses of 3 and 6 mg/day) study. Study RGH-MD-05 was a phase-3, randomized, double-blind, placebo-controlled, parallel-group, fixed/flexible-dose (two fixed dose ranges of 3-6 mg/day and 6-9 mg/day) study. Study RGH-MD-16, a dose-finding study, was a phase 2b, randomized, double-blind, placebo- and active-controlled, parallel-group, fixed-dose (three fixed doses of 1.5, 3, 4.5 mg/day) study. Study RGH-MD-03, a proof-of-concept study, was submitted as a supportive study. This study was a phase-2, randomized, double-blind, placebo-controlled, parallel-group study with cariprazine fixed-flexible doses, specifically two dose ranges of 1.5-4.5 mg/day and 6-12 mg/day. See Table 1 for more details.

In all the four studies, based on pre-specification, efficacy was evaluated based on the change from baseline to the 6-week endpoint score of PANSS total score (the primary efficacy variable) and CGI-S score (the key secondary efficacy variable).

2.1.3 Bipolar I Disorder Program

The cariprazine bipolar mania program included a total of three pivotal, 3-week, double-blind, placebo-controlled efficacy studies in adult patients with manic or mixed episodes associated with bipolar I disorder. The cariprazine doses ranged from 3 to 12 mg/day. Study RGH-MD-31 was a phase 2, randomized, double-blind, placebo-controlled, parallel-group, flexible-dose (a dose range of 3-12 mg/day) study. Study RGH-MD-32 was a phase-3, randomized, double-blind, placebo-controlled, parallel-group, flexible-dose (a dose range of 3-12 mg/day) study. Study RGH-MD-33 was a phase-3, randomized, double-blind, placebo-controlled, parallel-group,

⁴ See footnote 1.

⁵ See footnote 1.

fixed-flexible dose (two fixed dose ranges of 3-6 mg/day and 6-12 mg/day) study. See Table 1 for more details.

In all the three studies, based on pre-specification, efficacy was evaluated based on the change from baseline to the 3-week endpoint score of YMRS total score (the primary efficacy variable) and CGI-S score (the key secondary efficacy variable).

The treatment arms, the total number of randomized subjects of each study, the numbers of subjects randomized to each treatment arm and those of ITT subjects are listed in Table 1.

2.2 Data Sources

The sponsor's submission including the clinical study reports and datasets are available at the FDA server: <\\Cdseub5\evsprod\NDA204370\0000\m5>.

As shown in Table 1, in this original new drug application (NDA), the sponsor submitted seven randomized, double-blind, placebo-controlled, parallel-group studies for efficacy evaluation of cariprazine. Specifically, the submitted study data consist of (1) legacy data - SAS datasets of the sponsor-formatted clinical database, containing all data captured in the study case report forms as pre-specified, and (2) analysis datasets, which were generated from the clinical database by the sponsor.

The sponsor also submitted CDISC SDTM formatted clinical database as SAS datasets. The CDISC SDTM formatted SAS datasets should necessarily have equivalent information/contents to the SAS datasets of the sponsor-formatted clinical database. The CDISC SDTM datasets were generated from the sponsor's finalized clinical database.

The statistical evaluation of the sponsor's efficacy claims was based on (1) legacy data and (2) analysis datasets.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

This reviewer confirmed the sponsor's efficacy analyses based on the submitted analysis datasets that were generated from CDISC SDTM datasets. He verified the sponsor's efficacy results using the legacy data (data from the sponsor's clinical database).

Out of 220 unique sites that participated in cariprazine clinical studies of schizophrenia or bipolar mania, eight sites (5 US sites and 3 sites in Ukraine) were inspected by Office of Regulatory Affairs (ORA) as part of the pre-approval inspection activities for the cariprazine New Drug Application (NDA). The Office of Scientific Investigations (OSI) Investigator identified approximately 200 potentially unreported adverse events in the FDA Form 483 issued to one of the investigators who participated in the cariprazine clinical program. Unlike other GCP violations, this site was included in the efficacy analysis, as the GCP violation of this site

was found after the data was unblinded. The sponsor conducted efficacy sensitivity analyses for the pivotal studies that this investigator participated in. This site did not alter the efficacy conclusions.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

3.2.1.1 Schizophrenia studies

Study design:

Study RGH-MD-04: The duration of this study was 9 weeks: a no-drug washout period of up to 7 days followed by 6 weeks of double-blind treatment and a 2-week safety follow-up phase. At the end of the screening phase, patients meeting entry criteria for this study were randomized to 1 of 4 treatment groups in a 1:1:1:1 ratio: placebo, cariprazine 3 mg/day, cariprazine 6 mg/day, or aripiprazole 10 mg/day.

Study RGH-MD-05: The duration of this study was of 9 weeks: a no-drug washout phase of up to 7 days followed by 6 weeks of double-blind treatment and a 2-week safety follow-up phase. At the end of the screening period, patients meeting entry criteria for this study were randomized to 1 of 3 treatment groups in a 1:1:1 ratio: placebo, cariprazine 3-6 mg/day, or cariprazine 6-9 mg/day.

Study RGH-MD-16: The duration of this study was 9 weeks: it consisted of a no-drug washout period of up to 7 days, followed by 6 weeks of double-blind treatment and a 2-week safety follow-up period. At the end of the screening period, patients meeting entry criteria for this study were randomized to one of five treatment groups in a 1:1:1:1:1 ratio: placebo, cariprazine 1.5mg/day, cariprazine 3 mg/day, cariprazine 4.5 mg/day, or risperidone 4.0 mg/day.

In the three above studies, randomized patients remained hospitalized for a minimum of 4 weeks following the initiation of double-blind investigational product.

Study RGH-MD-03: The duration of this study was 10 weeks: 6 weeks of double-blind treatment and 4 weeks safety follow-up. A no-drug washout period of up to 7 days preceded randomization. At the end of the screening period, patients meeting entry criteria for this study were randomized (1:1:1 ratio) into one of three treatment groups: placebo, cariprazine 1.5-4.5 mg, or cariprazine 6-12 mg. Patients were hospitalized during the screening phase. Patients remained hospitalized for a minimum of 21 days following randomization and initiation of double-blind study drug.

Efficacy endpoints (Primary and Key Secondary): In these schizophrenia studies, to assess efficacy of cariprazine based on primary and key secondary efficacy variables (PANSS and CGI-S), subjects were designed to be exposed to placebo or cariprazine for a 6-week, double-blind phase, at the end of which efficacy comparisons were performed. In all studies, the change from baseline to Week 6 in PANSS total score (and CGI-S score) was used as the endpoint analysis variable.

3.2.1.2 Bipolar I disorder studies

Study design:

Study RGH-MD-31: The duration of this study was 5 weeks: 3 weeks of double-blind treatment and 2 weeks of safety follow-up. Patients started hospitalization during the screening phase. A no-drug washout period of up to 4 days preceded randomization. All patients meeting the eligibility criteria were randomized in a 1:1 ratio to cariprazine or placebo.

Study RGH-MD-32: The duration of this study was 6 weeks: a no-drug screening phase of up to 4 to 7 days, followed by 3 weeks of double-blind treatment and a 2-week safety follow-up phase. At the end of the screening phase, patients meeting the entry criteria for this study were randomized in a 1:1 ratio to placebo or cariprazine 3-12 mg/day.

Study RGH-MD-33: The duration of this study was 6 weeks: a no-drug screening phase of up to 4 to 7 days, followed by 3 weeks of double-blind treatment and a 2-week safety follow-up phase. At the end of the screening phase, patients meeting the entry criteria were randomized in a 1:1:1 ratio to placebo, cariprazine 3-6mg/day or cariprazine 6-12 mg/day.

In all bipolar mania studies, all patients started hospitalization during the screening phase. Randomized patients remained hospitalized for a minimum of 14 days following the initiation of double-blind medication during the double-blind treatment phase.

Efficacy endpoints (Primary and Key Secondary): In these bipolar mania studies, to assess efficacy of cariprazine based on primary and key secondary efficacy variables (YMRS and CGI-S), subjects were designed to be exposed to placebo or cariprazine for a 3-week, double-blind phase, at the end of which efficacy comparisons were performed. In all studies, the change from baseline to Week 3 in YMRS total score (and CGI-S score) was used as the endpoint analysis variable.

Note on dose escalation plan: In the fixed-flexible dose and flexible dose studies of both indications, the drug administration was planned so dose was titrated to the lower-end dose of the dose range, and then dose was optimized for the target dose (the upper-end dose of the dose range) by increasing dose by a pre-determined increment based on efficacy and safety. A dose increase decision was based on investigator's judgment or based on a rule (such as 50% improvement relative to the previous visit in the primary efficacy) by the investigator. See also the sponsor's *dose recommendation* of sections of *sponsor's efficacy conclusions* (Sections 3.2.4.1.1 and 3.2.4.2.1 of this review).

3.2.2 Statistical Methodologies

Analysis method:

(1) Primary analysis: In the phase-3 studies (MD-04 and MD-05 for schizophrenia and MD-32 and MD-33 for bipolar mania), the sponsor pre-specified an MMRM (mixed model with repeated measures) approach for the primary analysis method for both efficacy endpoints (PANSS or YMRS and CGI-S). In the phase-2 studies (MD-16 and MD-03 for schizophrenia and MD-31 for bipolar mania), the sponsor pre-specified an ANCOVA (LOCF) approach for the primary analysis method. Refer to Table 8, Table 10, Table 13 and Table 15.

(2) Sensitivity analysis: A sensitivity analysis was planned to assess the robustness of the primary analysis results to a possible violation of the assumption for the missing data mechanism. In the phase-3 studies (MD-04 and MD-05 for schizophrenia and MD-32 and MD-33 for bipolar mania), the sponsor pre-specified a PMM (pattern mixture model) approach and an ANCOVA (LOCF) approach for sensitivity analysis methods.

In the phase-2 studies (MD-16 and MD-03 for schizophrenia and MD-31 for bipolar mania), the sponsor pre-specified an MMRM approach for sensitivity analysis method.

MMRM: The pre-specified primary analysis used a mixed model with repeated measures (MMRM) with treatment group, pooled study center, visit, and treatment group by visit interaction as fixed effects and the baseline value and baseline value by visit interaction as covariates. An unstructured covariance matrix was used to model the covariance of within-patient scores. The Kenward-Roger approximation was used to estimate denominator degrees of freedom. This analysis was performed based on all post-baseline scores using only the observed cases without imputation of missing values. The missing at random (MAR) missing data mechanism is assumed in the MMRM approach. If this assumption is valid, missing data may be ignored in an application of MMRM to observed data.

ANCOVA (LOCF): The pre-specified primary analysis used an ANCOVA (LOCF) with treatment, pooled study center as factors, and the baseline score as a covariate.

PMM: A sensitivity analysis using a pattern-mixture model (PMM) based on non-future dependent missing value (NFDMV) restriction was performed to assess the robustness of the primary MMRM results to a possible violation of the missing-at-random assumption. With the proposed method⁶, missing values may be imputed to create a complete data (observed data plus imputed missing values). By analyzing the complete data with MMRM, it is feasible to assess the robustness of the MMRM approach under a specific MNAR (missing not at random) scenario.

A pattern-mixture model (PMM) approach is based on subjects classified by patterns indicating at which visit subjects had a missing observation in the first place and dropped out after the visit. Thus patterns for the PMM are defined by subjects' last visit with an observation. The non-future dependent missing value (NFDMV) restriction allows one to impute all missing observations using observed data under an assumption that missing data occur due to poor efficacy (one particular situation of MNAR). Using a multiple imputation (MI) technique, it is possible to construct an imputation distribution, for each pattern, from which missing data can be simulated. By including a shift parameter as a sensitivity parameter in missing data imputation distributions, the sponsor's proposed PMM with NFDMV facilitates sampling imputed missing observations. The dataset with missing values may be analyzed using the same model as the primary analysis for between-treatment group comparisons at Week 6. As in an MI application, the analysis was performed multiple times and the inference of this sensitivity analysis was based on the combined estimates using the standard multiple imputation technique. The value for the shift parameter was pre-specified.

⁶ The sponsor's proposed method is based on: *Kenward MG, Molenberghs G and Thijs H. (2003) Pattern-mixture models with proper time dependence. Biometrika, 90, 53-71*

Multiplicity adjustment method: The sponsor pre-specified a method for controlling study-wise type I error rate in every study. In the phase-3 studies (MD-04 and MD-05 for schizophrenia, and MD-33 for bipolar mania), the sponsor's proposed method was matched parallel gatekeeping (matched between the primary and the key secondary endpoints for each dose group).

Matched parallel gatekeeping: To control the overall type I error rate for multiple comparisons of the two cariprazine doses versus placebo across the primary and the key secondary hypotheses, the matched parallel gatekeeping procedure was planned⁷. In the following description, Study MD-04 is used as an example. The primary hypothesis family, F1, consists of two null hypotheses H_{11} and H_{12} for comparisons of cariprazine 3 mg and 6 mg, respectively, with placebo in regard to the primary efficacy parameter. That is, $F1 = \{H_{11}, H_{12}\}$. Similarly, we have the key secondary hypothesis family $F2 = \{H_{21}, H_{22}\}$, with H_{21} and H_{22} for comparisons of cariprazine 3 mg and 6 mg, respectively, with placebo in regard to the key secondary efficacy parameter. The family F1 serves as a gatekeeper for F2 such that F2 will be examined only when the gatekeeper F1 has been successfully passed (i.e., at least one of the hypotheses in the F1 family is rejected). The matched gatekeeper procedure utilizes the special logical relationship between the primary and the secondary parameters to enhance the power of statistical testing. The significance in the secondary endpoint for a dose level cannot be claimed unless its corresponding primary hypothesis is found significant.

Sequential testing: In Study MD-16, the sponsor pre-specified a sequential testing procedure (first PANSS and then CGI-S). Specifically, if the average effect of 3 mg and 4.5 mg is significant, then proceed to testing each of these doses. Otherwise, stop. If significant, then proceed to testing 1.5 mg. If all tests for PANSS are significant, tests for CGI-S will be performed in the same way as for PANSS. Otherwise, no test for CGI-S was performed. In Study MD-03, the sponsor pre-specified a sequential testing procedure (first PANSS and then CGI-S). Within each endpoint, an F test for overall effects was planned. If significant, the test procedure proceeds on to individual doses. In Studies MD-31 and MD-32, the sponsor proposed a sequential testing procedure (first YMRS and then CGI-S).

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

3.2.3.1 Patient disposition

By treatment group: For each indication, patient disposition data (numbers of randomized and ITT subjects, numbers of dropouts and dropout rates, and numbers of subjects who prematurely discontinued) are provided in Table 2 and Table 3.

In both indications, the ITT population did not have a substantial deviation in number of subjects from randomized subjects. Generally, the overall dropout rates for both indications were about 38.0% for the three schizophrenia pivotal studies and about 30% for the three bipolar mania pivotal studies. In bipolar studies, more cariprazine treated subjects discontinued due to adverse

⁷ The sponsor's proposed method is based on: Chen X, Luo X., Capizzi T. (2005), *The application of enhanced parallel gatekeeping strategies*, *Stat Med* 2005 May 15.24(9): 1385-97.

Table 2: Patient disposition (Schizophrenia studies)

Study	Treatment arms	#Randomized subjects (#ITT population)	#Completed subjects	#Dropouts vs. #Randomized subjects (dropout rate)	Reason for premature discontinuation			
					Adverse Event	Insufficient therapeutic response	Withdrawal of consent	Other
MD-04	Cariprazine 3 mg	155 (151)	104	51/155 (32.9%)	15	15	19	2
	Cariprazine 6 mg	157 (154)	97	60/157 (38.2%)	20	14	25	1
	Placebo	153 (149)	95	58/153 (37.9%)	17	20	17	4
MD-05	Cariprazine 3-6 mg	151 (147)	96	55/151 (36.4%)	14	12	25	4
	Cariprazine 6-9 mg	148 (147)	86	62/148 (41.9%)	13	13	32	4
	Placebo	147 (145)	88	59/147 (39.5%)	13	26	16	4
MD-16	Cariprazine 1.5 mg	145 (140)	90	55/145 (37.9%)	14	18	18	5
	Cariprazine 3 mg	147 (140)	96	50/147 (34.2%)	8	17	22	3
	Cariprazine 4.5 mg	148 (145)	98	49/148 (33.3%)	12	15	16	6
	Placebo	151 (148)	79	72/151 (47.7%)	22	33	14	3
MD-03	Cariprazine 1.5-4.5 mg	128 (122)	68	59/128 (46.5%)	12	25	15	7
	Cariprazine 6-12 mg	134 (129)	72	61/134 (45.9%)	10	21	24	4
	Placebo	130 (126)	69	60/130 (46.5%)	19	20	17	4

[Source: Reviewer's results]

Note: "Other" includes "did not meet inc/exc criteria", protocol violation" and "lost to follow-up".

Table 3: Patient disposition (Bipolar I disorder studies)

Study	Treatment arms	#Randomized subjects (#ITT population)	#Completed subjects	#Dropouts vs. #Randomized subjects (dropout rate)	Reason for premature discontinuation			
					Adverse Event	Insufficient therapeutic response	Withdrawal of consent	Other
MD-31	Cariprazine 3-12 mg	118 (118)	75	43/118 (36.4%)	17	11	13	2
	Placebo	120 (117)	73	45/120 (38.1%)	12	18	14	1
MD-32	Cariprazine 3-12 mg	158 (158)	108	50/158 (31.6%)	15	7	26	2
	Placebo	154 (152)	106	48/154 (31.2%)	11	16	17	4
MD-33	Cariprazine 3-6 mg	167 (165)	129	38/167 (22.8%)	15	2	18	3
	Cariprazine 6-12 mg	169 (167)	119	50/169 (29.6%)	25	5	17	3
	Placebo	161 (160)	122	39/161 (24.2%)	8	15	14	2

[Source: Reviewer's results]

Note: "Other" includes "did not meet inc/exc criteria", protocol violation" and "lost to follow-up".

events than placebo subjects, but this was not necessarily the case with schizophrenia studies. More subjects prematurely discontinued in placebo due to insufficient therapeutic response in both indications, but this was not the case in Study MD-03.

By treatment group and country: Approximately two-thirds of the patients in the Intent-to-Treat (ITT) Populations from Studies RGH-MD-04 and RGH-MD-16 and one-half of the patients in Study RGH-MD-05 were enrolled at centers outside of the United States (See also Table 4). All study centers in Study RGH-MD-03 were located in the United States. Overall, in schizophrenia pivotal studies, about 41% (597/1,466) of ITT subjects were US subjects⁸. In Studies RGH-MD-04 and RGH-MD-16, about 35% of ITT subjects were from the US. In Study RGH-MD-05, about 52% of ITT subjects were from the US (see Table 3).

Approximately 1,049 patients with schizophrenia received cariprazine at doses of 1.5 to 9 mg/day in these three studies. In schizophrenia pivotal studies, the dropout rates of countries with a relatively large sample size were not far from 40% (the numbers of dropouts by country are not included in the table).

Approximately 60%, 40%, and 50% of the patients in the Intent to Treat (ITT) Populations from Studies RGH-MD-31, RGH-MD-32, RGH-MD-33, respectively, were enrolled at centers in the United States (See also Table 5). Overall, in bipolar mania pivotal studies, about 52% (541/1,037) of ITT subjects were US subjects. In Study RGH-MD-31, about 62% of ITT subjects were from the US. In Study RGH-MD-32, about 42% of ITT subjects were from the US. And in Study RGH-MD-33, about 53% of ITT subjects were from the US (See Table 5).

A total of 612 patients were exposed to cariprazine in these three studies. In bipolar mania studies, the dropout rates of countries (excluding India) with a relatively large sample size were not very different from 30% (the numbers of dropouts by country are not included in the table).

Table 4: Number of ITT subjects by Country (Schizophrenia studies)

Study	RGH-MD-04			RGH-MD-16				RGH-MD-05			Total
	CP3	CP6	Placebo	CP1.5	CP3	CP4.5	Placebo	CP3-6	CP6-9	Placebo	
South Africa	-	-	-	-	-	-	-	2	1	1	4
Malaysia	-	-	-	3	4	5	4	-	-	-	16
Columbia	-	-	-	-	-	-	-	16	15	16	47
Romania	18	19	14	-	-	-	-	-	-	-	41
Ukraine	34	37	36	24	22	23	23	-	-	-	199
India	-	-	-	30	31	34	29	53	54	54	285
Russia	46	45	47	32	31	31	35	-	-	-	267
US	53	53	52	51	52	52	57	76	77	74	597
All	151	154	149	140	140	145	148	147	147	145	1466

[Source: Reviewer’s results]

Note: CP1.5 denotes “Cariprazine 1.5mg/day”, CP3 “Cariprazine 3mg/day”, CP4.5 “Cariprazine 4.5mg/day”, CP6 “Cariprazine 6mg/day”, CP3-6 “Cariprazine 3-6 mg/day”, CP6-9 “Cariprazine 6-9 mg/day”.

⁸ The subjects of active comparator arms (Studies RGH-MD-04 and RGH-MD-16) are excluded from the percentages.

Table 5: Number of ITT subjects by Country (Bipolar I disorder studies)

Study	RGH-MD-31		RGH-MD-32		RGH-MD-33			Total
	CP3-12	Placebo	CP3-12	Placebo	CP3-6	CP6-12	Placebo	
Croatia	-	-	-	-	4	4	5	13
Serbia	-	-	-	-	7	10	6	23
Romania	-	-	-	-	17	17	15	49
Ukraine	-	-	-	-	22	23	21	66
India	34	32	91	87	-	-	-	244
Russia	11	12	-	-	26	26	26	101
US	73	73	67	65	89	87	87	541
All	118	117	158	152	165	167	160	1037

[Source: Reviewer’s results]

Note: CP3-6 “Cariprazine 3-6 mg/day”, CP6-12 “Cariprazine 6-12 mg/day”, and CP3-12 “Cariprazine 3-12 mg/day”.

3.2.3.2 Demographics and baseline characteristics

The sponsor reported demographics and baseline characteristics in the clinical study reports. This reviewer found that the distributions of subjects were fairly comparable among randomized patient groups in terms of demographics (age, gender, race and ethnicity), baseline physical characteristics (weight, height, BMI), and medical and psychiatric history. The baseline characteristics data of age group, gender and race can be found in Section 4 of this review.

3.2.4 Primary Analysis - Results and Conclusions

3.2.4.1 Primary and key secondary efficacy in Schizophrenia pivotal/supportive studies

3.2.4.1.1 Sponsor’s efficacy analysis

Primary analysis for PANSS and CGI-S: The statistical test results that the sponsor obtained based on the pre-specified hypothesis testing procedure with an appropriate overall type I error rate control are summarized in Table 6. Based on these results, the sponsor concluded as follows:

- Both fixed doses (fixed doses of 3 mg and 6 mg) were statistically significant for the primary and key secondary efficacy endpoints (PANSS and CGI-S). (MD-04)
- Both dose ranges (fixed-flexible dose ranges of 3-6 mg and 6-9 mg) were statistically significant for the primary and key secondary efficacy endpoints (PANSS and CGI-S). (MD-05)
- Three fixed doses (fixed doses of 1.5 mg, 3 mg, and 4.5 mg) were statistically significant for the primary and key secondary efficacy endpoints (PANSS and CGI-S). (MD-16)

This reviewer confirmed the above conclusions. It is noted that in the sponsor’s efficacy analyses, the following study centers were excluded from the sponsor’s efficacy analyses due to GCP violations: Center IDs of 043 and 504 (Study MD-05)⁹. The study sites had no impact on the efficacy results.

⁹ See page 86 of the CSR.

Sponsor efficacy conclusions¹⁰:

- *Primary efficacy*: In three pivotal studies, cariprazine at doses of 1.5 to 9 mg administered once daily was significantly better than placebo at the 6-week endpoint in reduction of schizophrenia symptoms as measured by change in PANSS total score. A choice of a method for the primary analysis, MMRM or ANCOVA (LOCF), did not lead to a different conclusion (Table 8).
- *Dose recommendation*: The recommended starting dose is 1.5 mg once daily and can be increased to 3 mg on Day 2. Depending on response and tolerability, dose can be adjusted in increments of 1.5 mg or 3 mg (b) (4)
- *Dose response*: A dose-response relationship was established in each of the 3 pivotal studies, with higher doses demonstrating greater efficacy on the primary endpoint (Table 8).
- *Visit-wise efficacy profile*: The significant difference from placebo in the change from baseline to Week 6 in the PANSS total score was observed as early as Week 1 in the 3 pivotal studies and was sustained through the 6-week endpoint (Table 9 and Figure 1).
- *Key secondary efficacy*: The robustness of the efficacy results in each study was supported by the prospectively defined secondary parameter, change from baseline in CGI-S, which registered the assessment of the clinician under double-blind conditions. The significant improvement in CGI-S using both MMRM and LOCF analyses in the 3 pivotal studies established efficacy of cariprazine on reducing severity of illness.
- *Subgroup analysis*: The subgroup analyses by study showed that cariprazine is effective in men and in women and is not limited by age group, race (white and all other races), or geographic region.

Sponsor sensitivity analysis:

In the schizophrenia phase-3 studies (MD-04 and MD-05), the primary analysis method for both endpoints (PANSS and CGI-S) was based on an MMRM approach, and the sensitivity analysis method a PMM approach with a non-future dependent missing value (NFD MV) restriction. The results from the sponsor's sensitivity analysis suggest that the primary efficacy analysis conclusion was not sensitive to (not affected by) the choice of an assumption between MAR or the specified MNAR scenario. This result of robustness is meaningful, but does not generally hold for any MNAR situation. In other words, the result is only valid for the particular scenario set up for the sensitivity analysis.

This reviewer agrees to the sponsor's efficacy conclusions but those on dose recommendation and dose response.

¹⁰ The conclusions are found in the sponsor's submitted documents (Pages 26-29: *Section 2.5 Clinical Overview of Common Technical Document Summaries*).

Table 6: Sponsor’s efficacy conclusion based on pre-specified analysis (Schizophrenia)

Study and Analysis method	Pre-specified method for Multiplicity adjustment	Statistical test results	Conclusion
RGH-MD-04 [MMRM]	Matched parallel gatekeeping (matched between PANSS and CGI-S for each dose group)	Cariprazine 3 mg/day: adjusted p = 0.0044 for PANSS, adjusted p = 0.0044 for CGI-S	Both fixed doses for both endpoints were statistically significant at the 5% significance level.
		Cariprazine 6 mg/day: adjusted p < 0.0001 for PANSS, adjusted p < 0.0001 for CGI-S	
RGH-MD-05 [MMRM]	Matched parallel gatekeeping (matched between PANSS and CGI-S for each dose group)	Cariprazine 3-6 mg/day: adjusted p = 0.0029 for PANSS, adjusted p = 0.0115 for CGI-S	Both fixed-flexible doses for both endpoints were statistically significant at the 5% significance level.
		Cariprazine 6-9 mg/day: adjusted p < 0.0001 for PANSS, adjusted p = 0.0002 for CGI-S	
RGH-MD-16 [ANCOVA (LOCF)]	Sequential testing (first PANSS and then CGI-S) – If average effect of 3mg and 4.5mg is significant then proceed to testing each of these doses. Otherwise, stop. If significant, then proceed to testing 1.5mg. If all tests for PANSS are significant, tests for CGI-S will be performed in the same way as for PANSS. Otherwise, no test for CGI-S.	Cariprazine 1.5 mg/day: p = 0.0005 for PANSS, p = 0.004 for CGI-S	The average of 3mg and 4.5mg was statistically significant. Then the three fixed doses for PANSS (in the order of 4.5, 3, and 1.5) were also statistically significant at the 5% significance level. Then, proceeded to testing CGI-S. The results of CGI-S were positive as those of PANSS.
		Cariprazine 3 mg/day: p < 0.0001 for PANSS, p = 0.0003 for CGI-S	
		Cariprazine 4.5 mg/day: p < 0.0001 for PANSS, p < 0.0001 for CGI-S	
RGH-MD-03 [ANCOVA (LOCF)]	Sequential testing (first PANSS and then CGI-S). Within each endpoint, F test was used for overall effects before testing individual doses.	The overall (average) effect of the two cariprazine dose-ranges for PANSS was found “not statistically significant”. No further tests were performed. Note: Individual (unadjusted) p-values were as follows: (b) (4)	

[Source: Sponsor’s Clinical Study Report and Reviewer’s results]

Note: The listed adjusted p-values were obtained after adjusting multiplicity of the primary and key secondary endpoints (PANSS and CGI-S) and different dose groups.

3.2.4.1.2 Reviewer’s assessments

This reviewer disagrees with the following sponsor’s conclusions. There are two points to consider:

- The 9 mg daily dose may not be considered one of the effective doses, if “effective doses” are meant to be as effective *fixed* doses.
- The data may not suggest a strong indication of a dose-response relationship for the primary efficacy variable (PANSS).

The assessments presented in this section are based on MMRM analysis as specified by the sponsor (Table 7, Table 8, Table 9 and Table 10). The efficacy estimates (least square (LS) means) of each treatment group, the efficacy endpoint estimates (Placebo-subtracted LS mean of change from baseline to Week 6) obtained from MMRM and ANCOVA (LOCF) are provided in Table 8 (PANSS) and Table 10 (CGI-S), and plotted in Figure 1 (Placebo-subtracted LS mean of change from baseline) and Figure 2 (LS means of change from baseline).

Fixed-dose study results (1.5 mg, 3 mg, 4.5 mg, and 6 mg are effective):

For both primary and key secondary endpoints (PANSS and CGI-S), the statistical evidence of efficacy of all planned three daily doses, 1.5mg, 3mg and 4.5mg was established in a phase-2 dose finding study (MD-16). In addition, the phase-3, fixed dose study (MD-04) with daily fixed doses of 3mg and 6mg, designed based on the outcome of the phase-2 study, has shown statistically significant evidence of efficacy for both efficacy endpoints, with an appropriate overall type I error rate control (Table 6).

Efficacy of a fixed dose of 9 mg has not been established:

The sponsor concluded that cariprazine 1.5 to 9 mg/day has been shown to be effective for the treatment of schizophrenia. However, efficacy of a fixed daily dose of 9 mg was not assessed in any pivotal or supportive studies. No data is available to sufficiently assess whether or not the 9 mg is as effective as any other lower dose. (b) (4)

Studies that included subjects who received a daily dose of cariprazine 9 mg:

There is efficacy data of subjects exposed to 9 mg from Study MD-05. However, it may not be interpreted as suggesting efficacy of a daily fixed dose of 9 mg. Some subjects were exposed to the final (optimal) daily dose of 9 mg. In the group of a daily dose of 6-9 mg of Study MD-05, 92 subjects (62.2%) of 148 randomized subjects reached a daily dose of 9 mg after the investigators' Week 2 efficacy assessment based on PANSS and CGI-S, plus tolerability, and maintained the daily dose till the last treatment day for about 4 weeks. In short, a dose range of 6-9 mg was statistically shown to be efficacious, with 62.2% of the subjects exposed to a final daily dose of 9 mg. The efficacy results of this dose range, -9.9 (LS mean difference from placebo) and -25.9 (LS mean change from baseline), may appear to be relatively better than those of the dose range of 3-6 mg. Thus this might be interpreted as being suggestive of efficacy of a fixed daily dose of 9 mg (Table 8). However, the overall mean daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days was 6.55 mg/day, in the cariprazine 6-9 mg/day arm (Table 7). Thus efficacy outcome of the daily dose range of 6-9 mg in dose optimization may not be viewed as showing efficacy of a fixed daily dose of 9 mg. This reviewer also notes, in contrast, that in the fixed dose study (Study MD-04), subjects were designed to reach the target fixed doses (3 mg and 6 mg) at Day 4, and the great majority of subjects (over 90%) achieved the target dose in about a week after randomization. In this fixed dose study, the overall mean exposure was 5.63 mg/day, which is close to 6 mg/day. In the cariprazine 6-12 mg arm of Study MD-03, only 16 subjects (12.0%) out of 133 randomized subjects remained at the 9 mg daily dose till the last treatment day. The overall mean daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days, was 8.7 mg/day in cariprazine 6-12 mg/day (Table 7). The efficacy

outcome seems less efficacious when compared to the pivotal studies. That is, the LS mean difference from placebo was -3.8, and the LS mean change from baseline -16.8.

Efficacy of a fixed-flexible dose study:

In Study MD-05, efficacy of two daily dose ranges (3-6 mg and 6-9 mg) was statistically established for both primary and key secondary endpoints (PANSS and CGI-S), with an appropriate type I error rate control (Table 6). The LS mean change from baseline of two-dose ranges (3-6 mg and 6-9 mg), were -22.8 and -25.9, respectively. The LS mean difference from placebo of two-dose ranges (3-6 mg and 6-9 mg), were -6.8 and -9.9, respectively (Table 8).

A fixed-flexible dose range does not give a definitive conclusion on its dose responsive efficacy. The fixed-flexible dose studies MD-05 and MD-03 should not be considered comparable with the fixed dose study (MD-04), because of differences in overall mean daily exposure (Table 7). While in MD-04 over 90% of subjects reached the fixed doses, a much lower proportion of subjects reached the upper-end dose of the dose ranges in MD-03 and MD-05 (Table 7). The subject distribution in final daily doses taken for the last treatment day was 36 vs 111 subjects (23.8% vs 73.5%) for the lower-end dose of 3 mg and the upper-end dose of 6 mg, respectively, in cariprazine 3-6 mg/day (N=151). The subject distribution in final daily doses taken for the last treatment day was 47 vs 92 subjects (31.8% vs 62.2%) for the lower-end dose of 6mg and the upper-end dose of 9 mg, respectively, in cariprazine 6-9 mg/day (N=148). The overall mean daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days, was 4.22 mg/d in cariprazine 3-6 mg/day and 6.55 mg/day in cariprazine 6-9 mg/day¹¹. It may not be reasonable to interpret the efficacy results of the fixed-flexible dose study for discussing efficacy of fixed doses.

Differences in drug administration scheme in Studies MD-03 and MD-05:

It is noted that in Study MD-03, the investigators made a decision on dosing increase/decrease based on their judgment on efficacy and tolerability. This was different from Study MD-05, in which a decision on a dosing increase/decrease was based on efficacy response as measured with the primary efficacy variable and tolerability as judged by the investigator at post-titration visits. In the arm of cariprazine 6-12 mg/day of Study MD-03, a subject typically reached a daily dose of 9 mg at Day 7, and proceeded on to a 12 mg daily dose from Days 9-14. In the arm of cariprazine 1.5-4.5 mg/day of Study MD-03, a subject typically reached a daily dose of 4.5 mg at Day 5. Over 80% of subjects of each dose range (1.5-4.5 mg and 6-12 mg) reached the upper-end doses (4.5 mg and 12 mg, respectively); specifically, 85.8% of subjects had a final dose of 4.5 mg, and 83.5% a final dose of 12 mg. These proportions were much higher compared to Study MD-05. In this sense, Study MD-05 should not be regarded as a simple repetition of Study MD-03. The overall mean daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days, was 3.8 mg/d in cariprazine 1.5-4.5 mg/day and 8.7 mg/day in cariprazine 6-12 mg/day. The LS mean change from baseline of two dose-ranges (1.5-4.5 mg and 6-12 mg), were -18.0 and -16.8, respectively. The LS mean difference from placebo of two dose-

¹¹ See Table 12.1-2 (page 117) from Study report of RGH-MD-05.

ranges (1.5-4.5 mg and 6-12 mg), were -5.0 and -3.8, respectively. It may be difficult to interpret a dose response relationship from this study.

Dose response may not exist:

The sponsor concluded a dose-response relationship for efficacy, based on the observation that in each of the 3 pivotal studies, efficacy improved with increased cariprazine doses (See Section 3.2.4.1.1 of this review). However, efficacy estimates of fixed doses of 1.5 mg, 3 mg, 4.5 mg and 6 mg do not seem to suggest a dose-response relationship across the two fixed dose studies (see Table 8 and Figure 2). First of all, in the dose finding study (MD-16), the daily dose of 3 mg appears to be much the same as that of 1.5 mg, as LS mean change from baseline and LS mean difference from placebo of these two doses were very close to each other. Secondly, the observed differentials in LS mean difference from placebo between the doses are not large, and the observed LS mean change from baseline scores were within a somewhat narrow range (-23.8 to -20.2) (See Table 8, Figure 1 and Figure 2).

As the sponsor finds, within each of the two fixed dose studies, there may appear to be a dose-response relationship between 3 mg and 6 mg (MD-04) and between 3 mg and 4.5 mg (MD-16) in terms of LS mean change from baseline as well as LS mean difference from placebo. But it seems unclear whether or not these numerical differences are suggestive of a dose-response relationship.

Table 7: Overall mean daily exposure and Proportion of subjects who achieved target dose (Schizophrenia – flexible dosing studies)

Study	Treatment arms	LS mean change from baseline (MMRM – PANSS)	Proportion of subjects who reached the upper-end dose (of dose range) / the fixed target dose (%)	Overall mean daily exposure, rounded (mg/day)
MD-03	Cariprazine 1.5-4.5 mg/day	(b) (4)		
	Cariprazine 6-12 mg/day			
MD-05	Cariprazine 3-6 mg/day	-22.8	73.5	4.2
	Cariprazine 6-9 mg/day	-25.9	62.2	6.6

[Source: Sponsor’s Clinical Study Reports]

Table 8: MMRM and ANCOVA (LOCF) results for primary efficacy (PANSS total score)

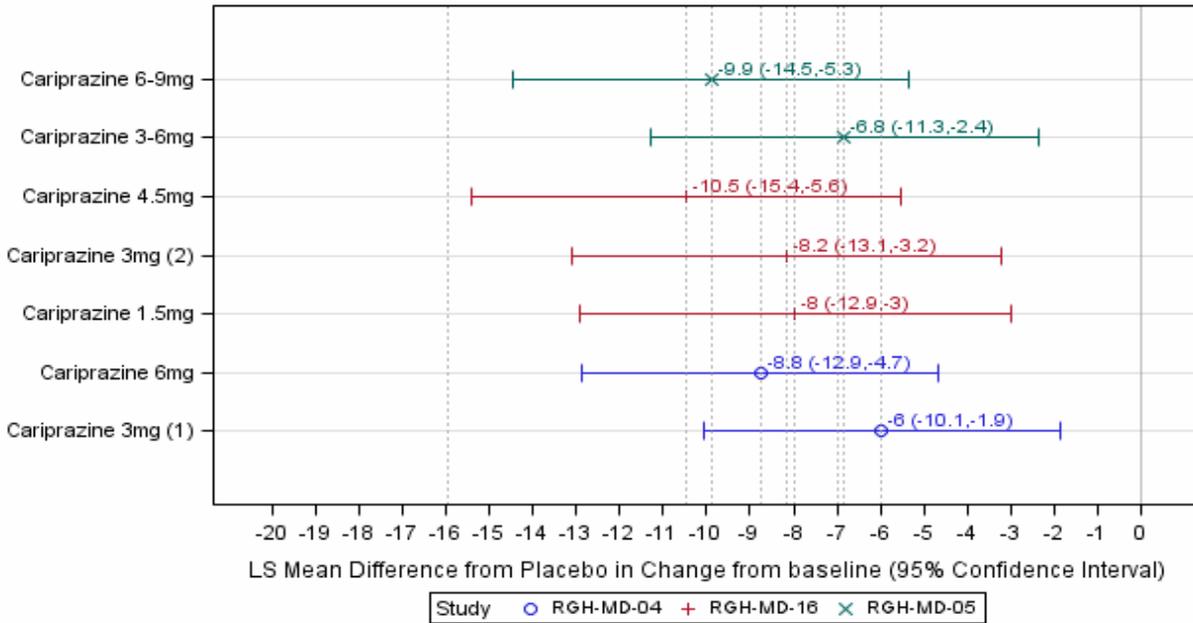
Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-04 (N=454)	Cariprazine 3 mg/day (N=151)	96.1 (8.7)	-20.2 (1.5)	-6.0	-16.4 (1.4)	-5.4
	Cariprazine 6 mg/day (N=154)	95.7 (9.4)	-23.0 (1.5)	-8.8	-18.9 (1.4)	-7.9
	Aripirazole 10 mg/day (N=150)	95.6 (9.0)	-21.2 (1.4)	-7.0	-18.8 (1.4)	-7.7
	Placebo (N=149)	96.5 (9.1)	-14.3 (1.5)	-	-11.0 (1.4)	-
MD-05 (N=439)	Cariprazine 3-6 mg/day (N=147)	96.3 (9.3)	-22.8 (1.6)	-6.8	-18.6 (1.5)	-6.3
	Cariprazine 6-9 mg/day (N=147)	96.3 (9.0)	-25.9 (1.7)	-9.9	-20.2 (1.5)	-8.0
	Placebo (N=145)	96.2 (9.3)	-16.0 (1.6)	-	-12.3 (1.5)	-
MD-16 (N=573)	Cariprazine 1.5 mg/day (N=140)	97.1 (9.1)	-21.3 (1.8)	-8.0	-19.4 (1.6)	-7.6
	Cariprazine 3 mg/day (N=140)	97.2 (8.7)	-21.5 (1.7)	-8.2	-20.7 (1.6)	-8.9
	Cariprazine 4.5 mg/day (N=145)	96.7 (9.0)	-23.8 (1.7)	-10.5	-22.3 (1.6)	-10.5
	Risperidone 4 mg/day (N=138)	98.1 (9.4)	-29.3 (1.7)	-16.0	-27.0 (1.6)	-15.2
	Placebo (N=148)	97.3 (9.2)	-13.3 (1.8)	-	-11.8 (1.5)	-
MD-03 (N=377)	Cariprazine 1.5-4.5 mg/day (N=122)					
	Cariprazine 6-12 mg/day (N=129)					
	Placebo (N=126)					

(b) (4)

[Source: Sponsor's Clinical Study Reports

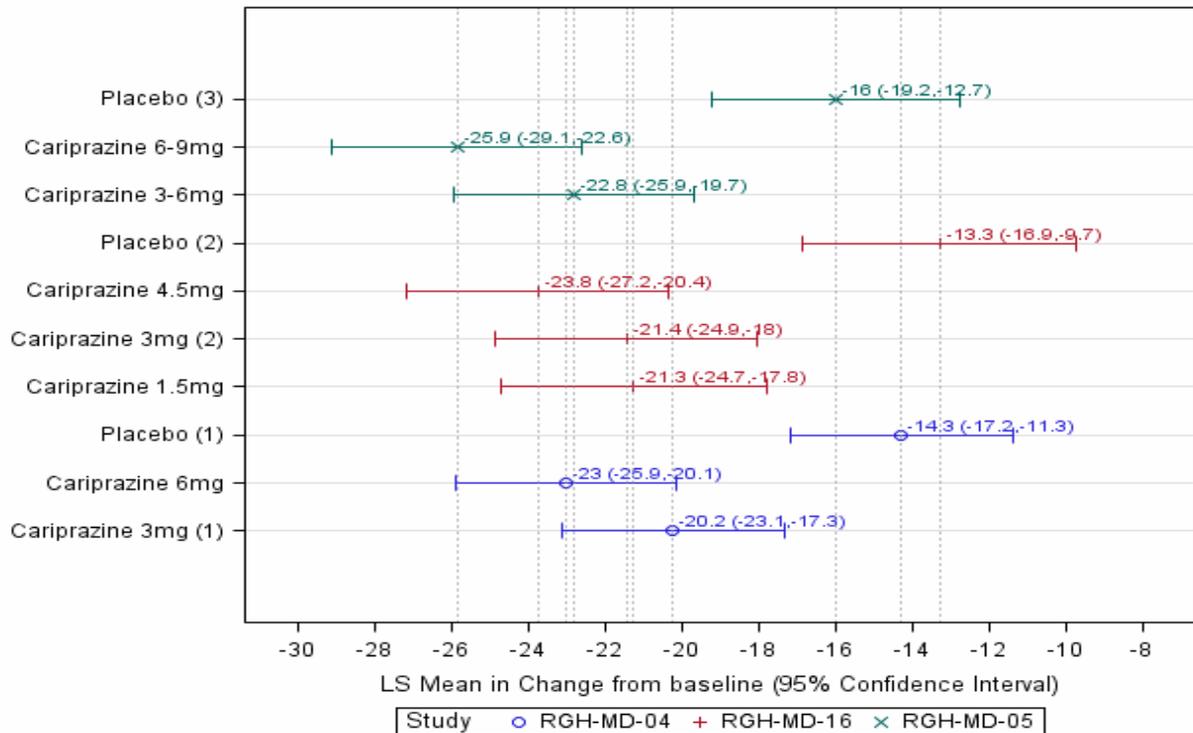
Note: (1) The bolded results are from pre-specified primary analysis. (2) Individual unadjusted 95% confidence intervals for mean difference from placebo are provided in Table 9 and Figure 1. (3) Adjusted *p* values for the primary endpoints (Studies MD-04, MD-05 and MD-16) are given in Table 6.

Figure 1: Endpoint LS mean difference from placebo and 95% confidence interval (Schizophrenia primary analysis (PANSS))



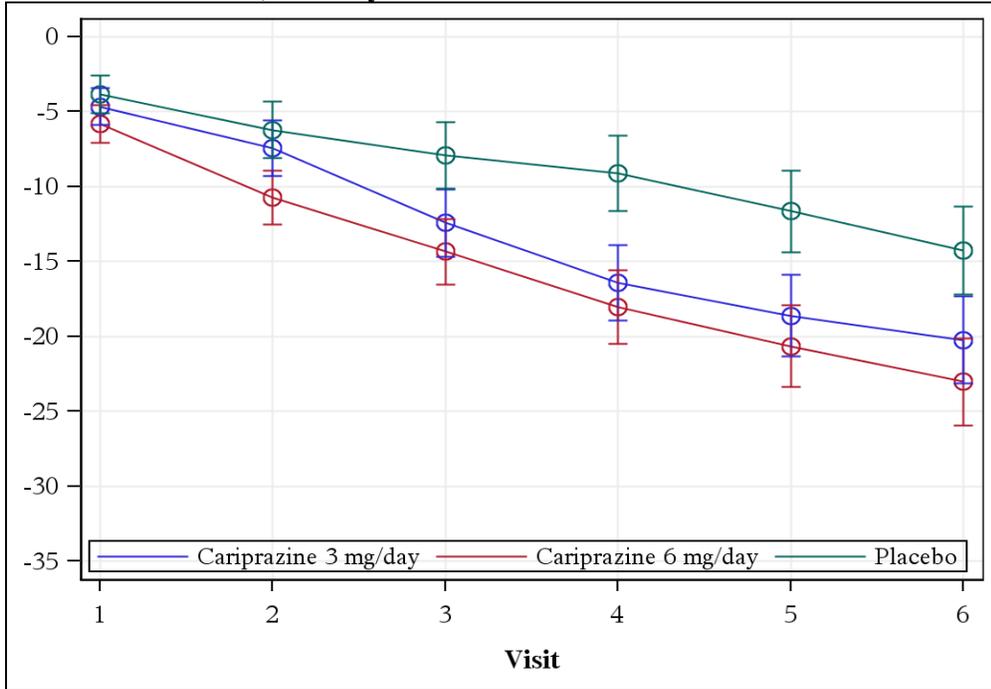
[Source: Reviewer's results]

Figure 2: Endpoint LS mean estimates and 95% confidence intervals (Schizophrenia primary analysis (PANSS))



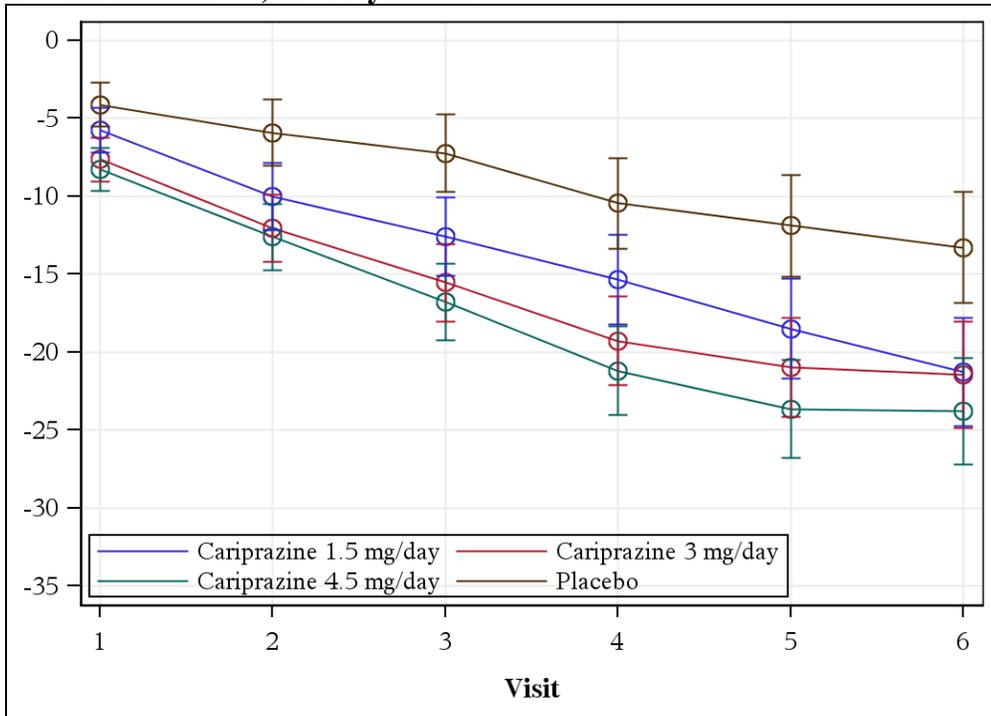
[Source: Reviewer's results]

Figure 3: Visit-wise LS mean estimates and 95% Confidence Intervals (Change from Baseline in PANSS) – Study MD-04



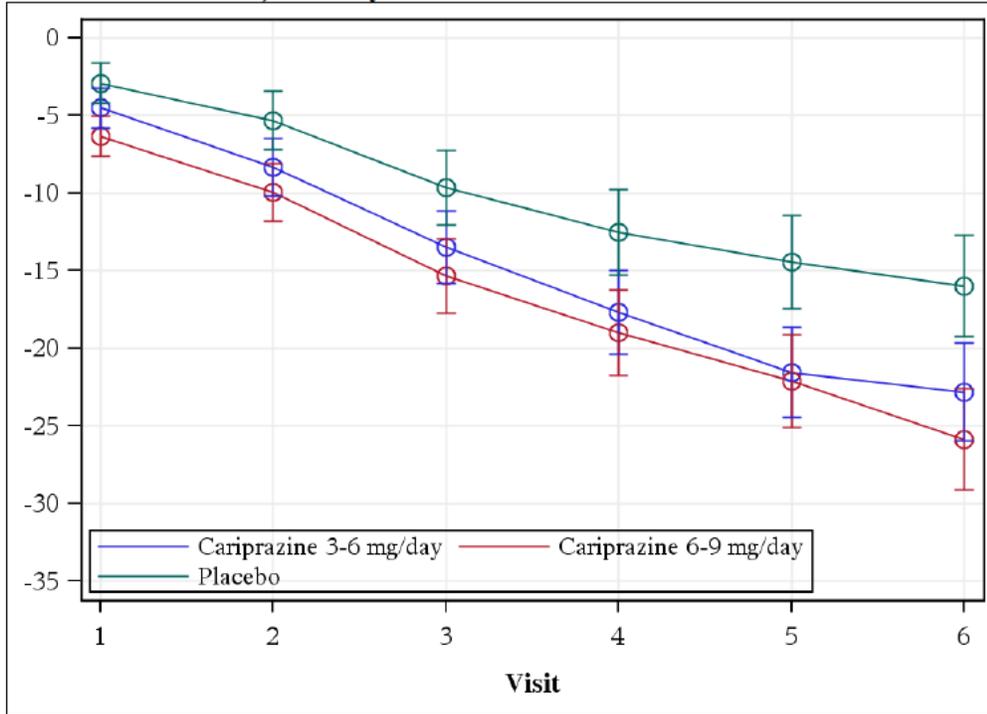
[Source: Reviewer's results]

Figure 4: Visit-wise LS mean estimates and 95% Confidence Intervals (Change from Baseline in PANSS) – Study MD-16



[Source: Reviewer's results]

Figure 5: Visit-wise LS mean estimates and 95% Confidence Intervals (Change from Baseline in PANSS) – Study MD-05



[Source: Reviewer's results]

Table 9: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in PANSS) – Schizophrenia pivotal studies

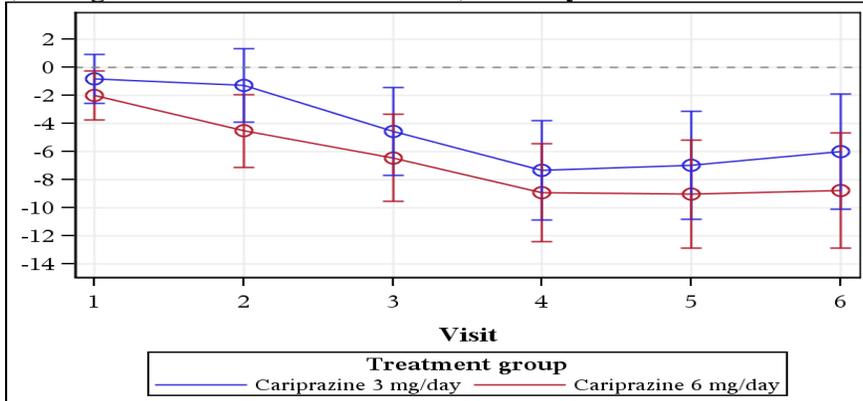
Study	Treatment arm	Visit	LS Mean Difference	Standard Error	95% Confidence Interval
RGH-MD-04	Cariprazine 3 mg/day	1	-0.8	0.9	(-2.6, 0.9)
		2	-1.2	1.3	(-3.9, 1.4)
		3	-4.5	1.6	(-7.7, -1.4)
		4	-7.3	1.8	(-10.8, -3.8)
		5	-7.0	2.0	(-10.8, -3.1)
		6	-6.0	2.1	(-10.1, -1.9)
	Cariprazine 6 mg/day	1	-2.0	0.9	(-3.7, -0.2)
		2	-4.5	1.3	(-7.1, -1.9)
		3	-6.4	1.6	(-9.5, -3.3)
		4	-8.9	1.8	(-12.4, -5.4)
		5	-9.0	2.0	(-12.8, -5.2)
		6	-8.8	2.1	(-12.9, -4.7)

Study	Treatment arm	Visit	LS Mean Difference	Standard Error	95% Confidence Interval	
RGH-MD-16	Cariprazine 1.5 mg/day	1	-1.7	1.0	(-3.6, 0.3)	
		2	-4.1	1.5	(-7.1, -1.1)	
		3	-5.4	1.8	(-8.9, -1.9)	
		4	-4.9	2.1	(-8.9, -0.8)	
		5	-6.6	2.3	(-11.2, -2.0)	
		6	-8.0	2.5	(-12.9, -3.0)	
	Cariprazine 3 mg/day	1	-3.5	1.0	(-5.5, -1.6)	
		2	-6.1	1.5	(-9.1, -3.1)	
		3	-8.3	1.8	(-11.8, -4.8)	
		4	-8.8	2.1	(-12.9, -4.8)	
		5	-9.1	2.3	(-13.6, -4.5)	
		6	-8.2	2.5	(-13.1, -3.2)	
	Cariprazine 4.5 mg/day	1	-4.2	1.0	(-6.1, -2.2)	
		2	-6.7	1.5	(-9.7, -3.7)	
		3	-9.5	1.8	(-13.0, -6.1)	
		4	-10.7	2.1	(-14.8, -6.7)	
		5	-11.8	2.3	(-16.3, -7.2)	
		6	-10.5	2.5	(-15.4, -5.6)	
	RGH-MD-05	Cariprazine 3-6 mg/day	1	-1.6	0.9	(-3.3, 0.1)
			2	-3.0	1.3	(-5.6, -0.4)
			3	-3.8	1.7	(-7.1, -0.5)
			4	-5.2	1.9	(-9.0, -1.3)
			5	-7.1	2.1	(-11.3, -3.0)
			6	-6.8	2.3	(-11.3, -2.4)
Cariprazine 6-9 mg/day		1	-3.4	0.9	(-5.1, -1.7)	
		2	-4.6	1.3	(-7.2, -2.0)	
		3	-5.7	1.7	(-9.0, -2.3)	
		4	-6.5	2.0	(-10.3, -2.6)	
		5	-7.7	2.1	(-11.9, -3.5)	
		6	-9.9	2.3	(-14.5, -5.3)	

[Source: Reviewer's results]

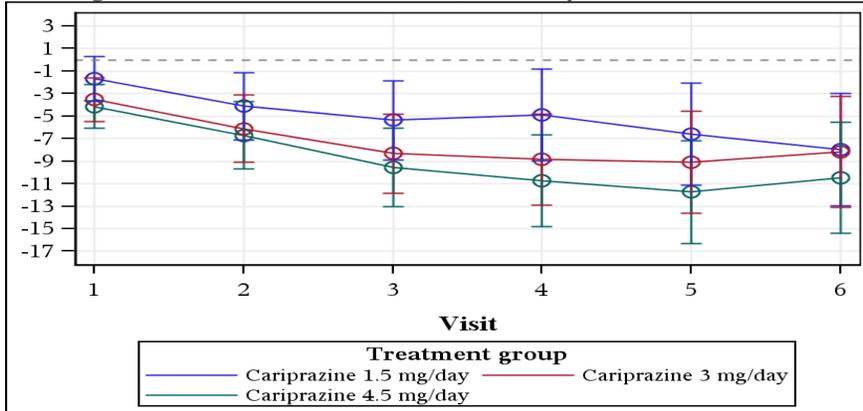
Note: Visit 6 was the Week 6 endpoint for all studies list in the table.

Figure 6: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in PANSS) – Study MD-04



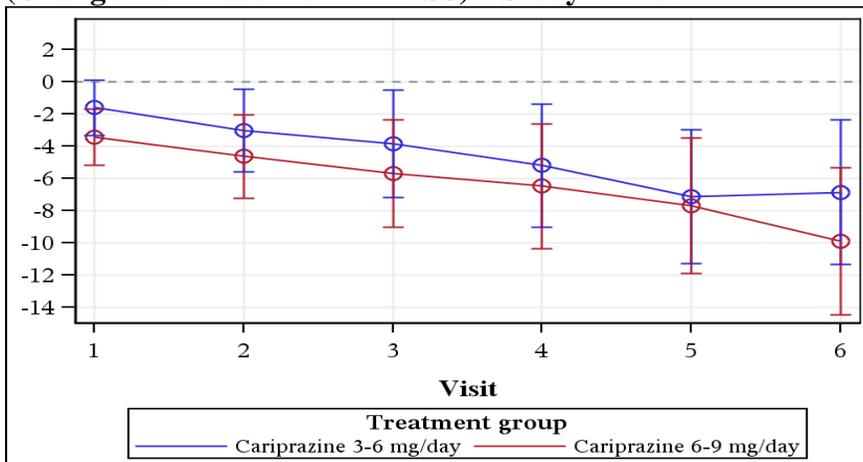
[Source: Reviewer's results]

Figure 7: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in PANSS) – Study MD-16



[Source: Reviewer's results]

Figure 8: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in PANSS) – Study MD-05



[Source: Reviewer's results]

Table 10: MMRM and ANCOVA (LOCF) results for key secondary efficacy (CGI-S)

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline Score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-04 (N=454*)	Cariprazine 3 mg/day (N=151)	4.9 (0.6)	-1.4 (0.1)	-0.4	-1.1 (0.1)	-0.4
	Cariprazine 6 mg/day (N=154)	4.8 (0.6)	-1.5 (0.1)	-0.5	-1.2 (0.1)	-0.5
	Aripiprazole 10 mg/day (N=150)	4.8 (0.6)	-1.4 (0.1)	-0.4	-1.2 (0.1)	-0.5
	Placebo (N=149)	4.8 (0.6)	-1.0 (0.1)	-	-0.7 (0.1)	-
MD-05 (N=439)	Cariprazine 3-6 mg/day (N=147)	4.8 (0.7)	-1.4 (0.1)	-0.3	-1.0 (0.1)	-0.3
	Cariprazine 6-9 mg/day (N=147)	4.9 (0.7)	-1.6 (0.1)	-0.5	-1.2 (0.1)	-0.4
	Placebo (N=145)	4.9 (0.7)	-1.0 (0.1)	-	-0.7 (0.1)	-
MD-16 (N=573*)	Cariprazine 1.5 mg/day (N=140)	4.7 (1.2)	-1.2 (0.1)	-0.3	-0.9 (0.1)	-0.4
	Cariprazine 3 mg/day (N=140)	4.9 (1.2)	-1.2 (0.1)	-0.3	-1.1 (0.1)	-0.5
	Cariprazine 4.5 mg/day (N=145)	4.8 (1.2)	-1.5 (0.1)	-0.6	-1.2 (0.1)	-0.6
	Risperidone 4 mg/day (N=138)	4.8 (1.2)	-1.6 (0.1)	-0.7	-1.4 (0.1)	-0.8
	Placebo (N=148)	4.9 (1.2)	-0.9 (0.1)	-	-0.6 (0.1)	-
MD-03 (N=377)	Cariprazine 1.5-4.5 mg/day (N=122)					
	Cariprazine 6-12 mg/day (N=129)					
	Placebo (N=126)					

(b) (4)

[Source: Sponsor's Clinical Study Reports and Reviewer's results]

Note: (1) The bolded results are from pre-specified primary analysis. (2) Adjusted *p* values for the key secondary endpoints are given in Table 6.

3.2.4.2 Primary and key secondary efficacy in Bipolar I disorder pivotal studies

3.2.4.2.1 Sponsor's analysis results

Primary analysis for YMRS and CGI-S: The statistical test results that the sponsor obtained based on the pre-specified hypothesis testing procedure with an appropriate overall type I error rate control are summarized in Table 11. Based on these results, the sponsor concluded as follows:

- In both studies (MD-31 and MD-32), cariprazine flexible dose (3-12 mg/day) was statistically significant at the 5% significance level sequentially for the primary and key secondary efficacy endpoints (YMRS and CGI-S).

- Both fixed-flexible dose ranges (3-6 mg/day and 6-12 mg/day) were statistically significant at the 5% significance level for the primary and key secondary efficacy endpoints (YMRS and CGI-S). (MD-33)

This reviewer confirms and agrees to the above conclusions. It is noted that the following study centers were excluded from the sponsor’s efficacy analyses due to GCP violations: Center IDs of 004 (Study MD-33)¹² and Center IDs of 105 (Study MD-32)¹³. The study sites had almost no impact on the efficacy results.

Table 11: Sponsor’s efficacy conclusion based on pre-specified analysis (Bipolar I Disorder)

Study and Analysis method	Pre-specified method for Multiplicity adjustment	Statistical test results	Conclusion
RGH-MD-31 [ANCOVA (LOCF)]	Sequential testing (first YMRS and then CGI-S)	Cariprazine 3-12 mg/day: p < 0.0001 for YMRS, p = 0.0001 for CGI-S	In both studies, cariprazine flexible dose (3-12 mg/day) was statistically significant at the 5% significance level sequentially for YMRS and then CGI-S.
RGH-MD-32 [MMRM]	Sequential testing (first YMRS and then CGI-S)	Cariprazine 3-12 mg/day: p = 0.0004 for YMRS, p = 0.0027 for CGI-S	
RGH-MD-33 [MMRM]	Matched parallel gatekeeping (matched between YMRS and CGI-S for each flexible-fixed dose group)	Cariprazine 3-6 mg/day: adjusted p < 0.001 for YMRS, adjusted p < 0.001 for CGI-S	Both cariprazine fixed-flexible doses (3-6 mg/day and 6-12 mg/day) were statistically significant at the 5% significance level for both YMRS and CGI-S.
		Cariprazine 6-12 mg/day: adjusted p < 0.001 for YMRS, adjusted p < 0.001 for CGI-S	

[Source: Sponsor’s Clinical Study Report and Reviewer’s results]

Note: The listed adjusted p-values were obtained after adjusting multiplicity of the primary and key secondary endpoints (YMRS and CGI-S) and different dose groups.

Sponsor efficacy conclusions¹⁴:

- *Primary efficacy:* In three pivotal studies, cariprazine at doses of 3 to 12 mg administered once daily was significantly better than placebo at the 3-week endpoint in reduction of manic symptoms as measured by change in YMRS total score. A choice of a method for the primary analysis, MMRM or ANCOVA (LOCF), did not lead to a different conclusion (Table 13).
- *Dose recommendation:* The recommended and target dose is 3 mg to (b) (4) mg once daily. The starting dose is 1.5 mg/day, increasing to Day 2. Depending on individual patient response and tolerance, the dose can be increased in increments of 1.5 mg or 3 mg (b) (4).
- *Relative efficacy of two dose ranges:* Study MD-33 did not demonstrate increased benefit in the 6-12 mg group relative to that in the 3-6 mg group (Table 13, Figure 9 and Figure 10).

¹² See page 83 of the CSR.

¹³ See page 76 of the CSR.

¹⁴ The conclusions are found in the sponsor’s submitted documents (Pages 30-34: *Section 2.5 Clinical Overview of Common Technical Document Summaries*).

- *Visit-wise efficacy profile:* The significant difference from placebo in the change from baseline to Week 3 in the YMRS total score was observed within 4 to 7 days in the 3 studies and persisted up to the 3-week endpoint using MMRM or LOCF (Table 14).
- *Key secondary efficacy:* The robustness of the efficacy results in each study was supported by the prospectively defined secondary parameter, change from baseline in CGI-S, which registered the assessment of the clinician under double-blind conditions. The significant improvement in CGI-S using both MMRM and LOCF analyses in the 3 pivotal studies established efficacy of cariprazine on reducing severity of illness. Clinically relevant improvement with cariprazine is further evidenced by significant improvements on the CGI-I score across all 3 studies.
- *Subgroup analysis:* The subgroup analyses by study showed that cariprazine is effective in men and in women and is not limited by age group, race (white and all other races), or geographic region.

Sponsor sensitivity analysis:

In the bipolar mania phase-3 studies, (MD-32 and MD-33) the primary analysis method for both endpoints (YMRS and CGI-S) was based on an MMRM approach, and the sensitivity analysis method a PMM approach with an NFDMMV restriction. The results from the sponsor's sensitivity analysis suggest that the primary efficacy analysis conclusion was not sensitive to (not affected by) the choice of an assumption between MAR or the specified MNAR scenario. This result of robustness is meaningful, but does not generally hold for any MNAR situation. In other words, the result is only valid for the particular scenario set up for the sensitivity analysis.

This reviewer has no disagreement on the sponsor's conclusions.

3.2.4.2.2 Reviewer's assessments

The assessments presented in this section are based on MMRM analysis as specified by the sponsor (Table 12, Table 13, Table 14 and Table 15). The efficacy estimates (least square (LS) means) of each treatment group, the efficacy endpoint estimates (Placebo subtracted LS mean of change from baseline to Week 6) obtained from MMRM and ANCOVA (LOCF) applications are provided in Table 13 (YMRS) and Table 15 (CGI-S), and plotted in Figure 9 (Placebo subtracted LS mean of change from baseline) and Figure 10 (LS means of change from baseline).

For both primary and key secondary endpoints (YMRS and CGI-S), the statistical evidence of efficacy of a dose range of 3-12 mg was established in a phase-2, flexible dose study (RGH-MD-31). The phase-3, flexible dose study (MD-32) with the same dose range as Study MD-31 has also shown statistically significant evidence of efficacy for both endpoints, with an appropriate overall type I error rate control (Table 11).

It is noted that the designs of Studies MD-31 and MD-32 were similar in many ways, but that the dose increase/decrease schemes were not the same, as described below.

In Study MD-31, the investigator decided to maintain, increase or decrease a subject's dose based on his/her judgment on efficacy and tolerability. The subject distribution in final daily doses (6 mg, 9 mg and 12 mg) of cariprazine 3-12 mg/day taken for the last treatment day was 15, 20 and 78 subjects (12.7%, 16.9% and 66.1%, Total N=118), respectively. The overall mean

daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days, was 8.8 mg/day¹⁵.

In Study MD-32, on Day 2 (Visit 2), the investigator used his/her judgment to decide if a subject should remain at a daily dose of 3 mg or increase to a daily dose of 6 mg, but on the succeeding visits (Visit 3 through 6), the investigator used a criterion of adequate response (YMRS \geq 50% improvement relative to the previous visit) to decide on a dose increase. The subject distribution in final daily doses (6 mg, 9 mg and 12 mg) of cariprazine 3-12 mg/day taken for the last treatment day was 34, 47 and 62 subjects (21.5%, 29.7% and 39.2%, Total N=158), respectively. The overall mean daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days, was 7.49 mg/day¹⁶.

In Study MD-33, efficacy of two daily dose ranges (3-6 mg and 6-12 mg) was statistically established for both primary and key secondary endpoints (YMRS and CGI-S), with an appropriate type I error rate control. However, this study did not demonstrate increased benefit in the 6- to 12-mg group relative to that in the 3- to 6-mg group. The subject distribution in final daily doses of cariprazine 3-6 mg/day (the lower-end dose of 3 mg, the middle dose of 4.5 mg and the upper-end dose of 6 mg) taken for the last treatment day was 17, 26 and 123 subjects (10.2%, 15.6% and 73.7%, Total N=167). Similarly, the subject distribution in final daily doses of cariprazine 6-12 mg/day (the lower-end dose of 6 mg, the middle dose of 9 mg and the upper-end dose of 12 mg) taken for the last treatment day was 19, 28 and 118 subjects (11.2%, 16.6% and 69.8%, Total N=169). The overall mean daily exposure, calculated as total dose divided by the total duration of the double-blind treatment phase in days, was 4.81 mg/day in cariprazine 3-6 mg/day and 9.05 mg/day in cariprazine 6-12 mg/day¹⁷. There was a difference (4.8 mg/day vs. 9.1 mg/day) in overall mean daily exposure between cariprazine 3-6 mg/day and cariprazine 6-12 mg/day, but their efficacy results (LS mean change from baseline) were much the same (Table 12 and Table 13).

The LS mean change from baseline score of the four cariprazine treatment arms and the overall mean daily exposure given in the above paragraphs are listed in Table 12. Among the four cariprazine treatment arms (two 3-12 mg/day arms, one 3-6 mg/day and one 6-12 mg/day), differentials in LS mean change from baseline do not suggest a dose-response relationship when the overall mean daily exposure is considered as “dose”. The same intuition applies in the relation between the proportion of subjects who reached the upper-end dose and LS mean change from baseline.

¹⁵ See Table 10.5.1-1 (page 65) from Study report of RGH-MD-31.

¹⁶ See Table 12.1-2 (page 97) from Study report of RGH-MD-32.

¹⁷ See Table 12.1-2 (page 109) from Study report of RGH-MD-33.

Table 12: Overall mean daily exposure and Proportion of subjects who achieved target dose (Bipolar mania flexible dosing studies)

Study	Treatment arms	LS mean change from baseline (MMRM – YMRS)	Proportion of subjects who reached the upper-end dose (%)	Overall mean daily exposure, rounded (mg/day)
MD-31	Cariprazine 3-12 mg/day	-15.5	66.1	8.8
MD-32	Cariprazine 3-12 mg/day	-19.6	39.2	7.5
MD-33	Cariprazine 3-6 mg/day	-18.6	73.7	4.8
	Cariprazine 6-12 mg/day	-18.5	69.8	9.1

[Source: Sponsor’s Clinical Study Reports]

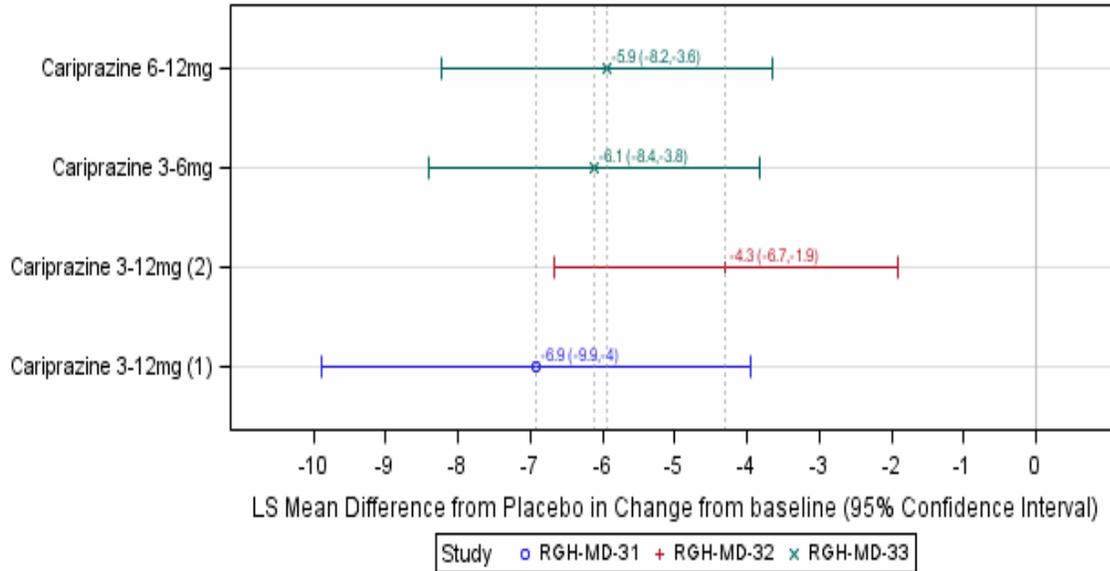
Table 13: MMRM and ANCOVA (LOCF) results for primary efficacy (YMRS total score)

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-31 (N=235)	Cariprazine 3-12 mg/day (N=118)	30.6 (5.0)	-15.5 (1.1)	-7.0	-13.3 (1.1)	-6.1
	Placebo (N=117)	30.2 (5.2)	-8.5 (1.1)	-	-7.2 (1.1)	-
MD-32 (N=310)	Cariprazine 3-12 mg/day (N=158)	32.3 (5.8)	-19.6 (0.9)	-4.3	-17.3 (0.9)	-4.3
	Placebo (N=152)	32.1 (5.6)	-15.3 (0.9)	-	-13.0 (0.9)	-
MD-33 (N=492)	Cariprazine 3-6 mg/day (N=165)	33.2 (5.6)	-18.6 (0.8)	-6.1	-17.9 (0.8)	-5.8
	Cariprazine 6-12 mg/day (N=167)	33.0 (4.7)	-18.5 (0.8)	-5.9	-17.4 (0.8)	-5.3
	Placebo (N=160)	32.6 (5.8)	-12.5 (0.8)	-	-12.1 (0.8)	-

[Source: Sponsor’s Clinical Study Reports and Reviewer’s results]

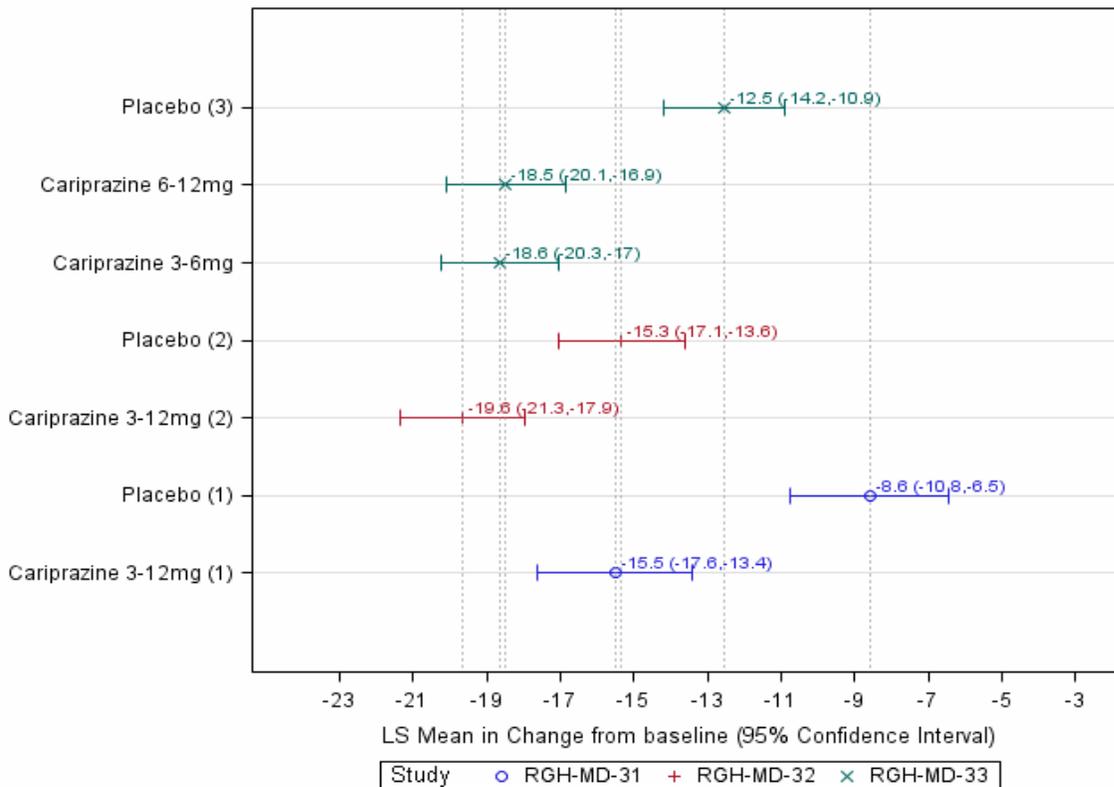
Note: (1) The bolded results are from pre-specified primary analysis. (2) In MMRM analysis of Study RGH-MD-31, an interaction term of Baseline score and Visit was not included in the model. This reviewer confirmed the results are almost identical when this interaction term is included. (3) Individual unadjusted 95% confidence intervals for mean difference from placebo are provided in Table 14 and Figure 3. (In the MMRM analysis of Study MD-31, this reviewer used a model that includes an interaction term of Baseline score and Visit, and thus the LS mean difference from placebo was different from that of Table 13. The difference was negligible.) (4) Adjusted *p* values for the primary endpoints are given in Table 11.

Figure 9: Endpoint LS mean difference from placebo and 95% confidence intervals (Bipolar mania primary analysis (YMRS))



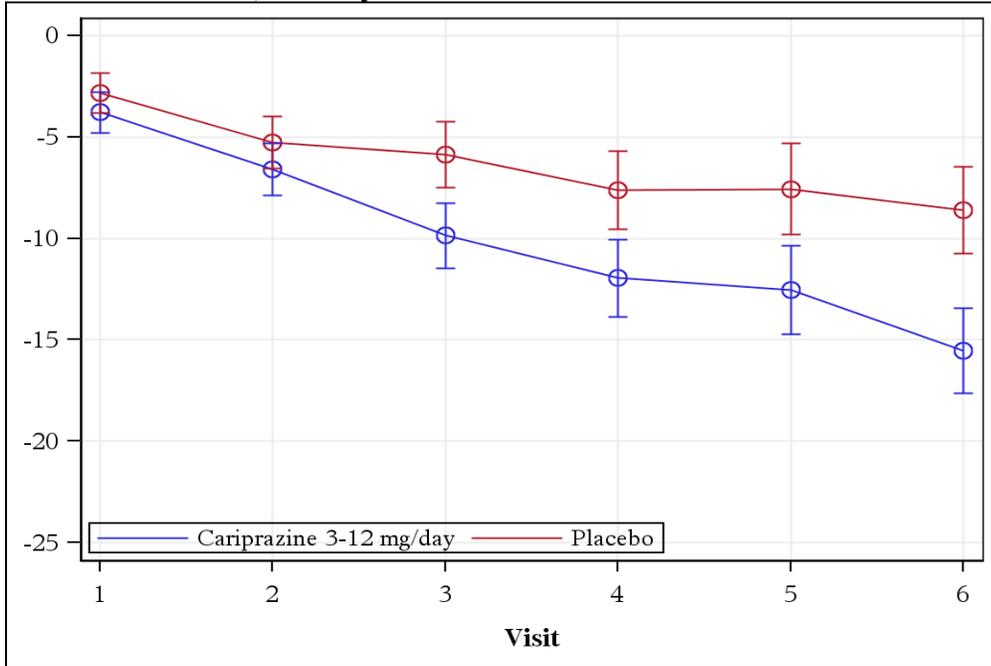
[Source: Reviewer's results]

Figure 10: Endpoint LS mean estimates and 95% confidence intervals (Bipolar mania primary analysis (YMRS))



[Source: Reviewer's results]

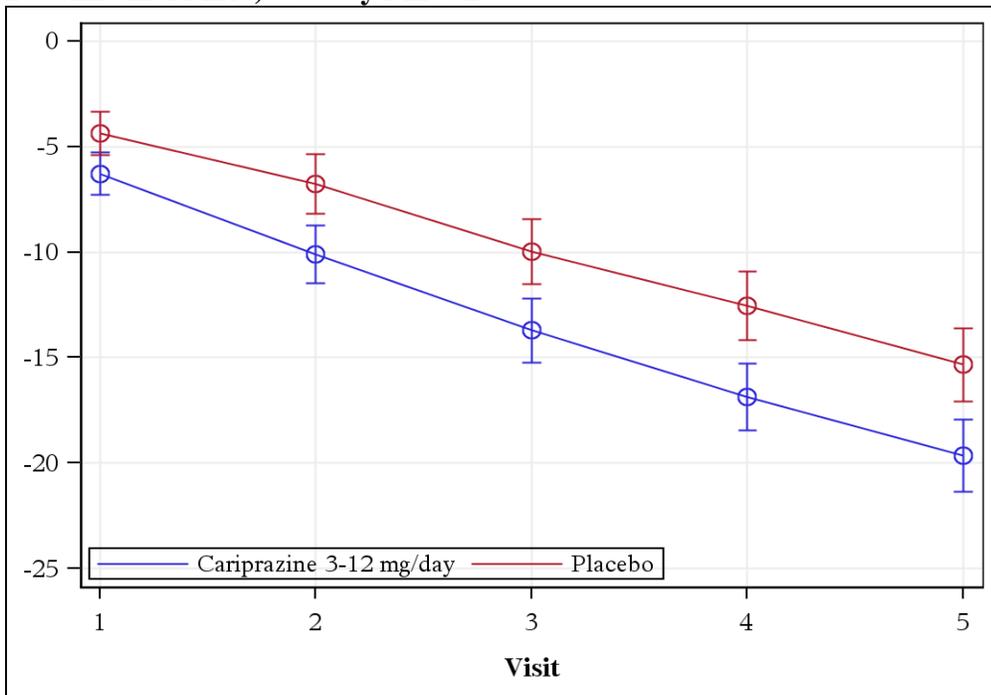
Figure 11: Visit-wise LS mean estimates and 95% Confidence Intervals (Change from Baseline in YMRS) – Study MD-31



[Source: Reviewer's results]

Note: Visit 6 is the 3-week endpoint.

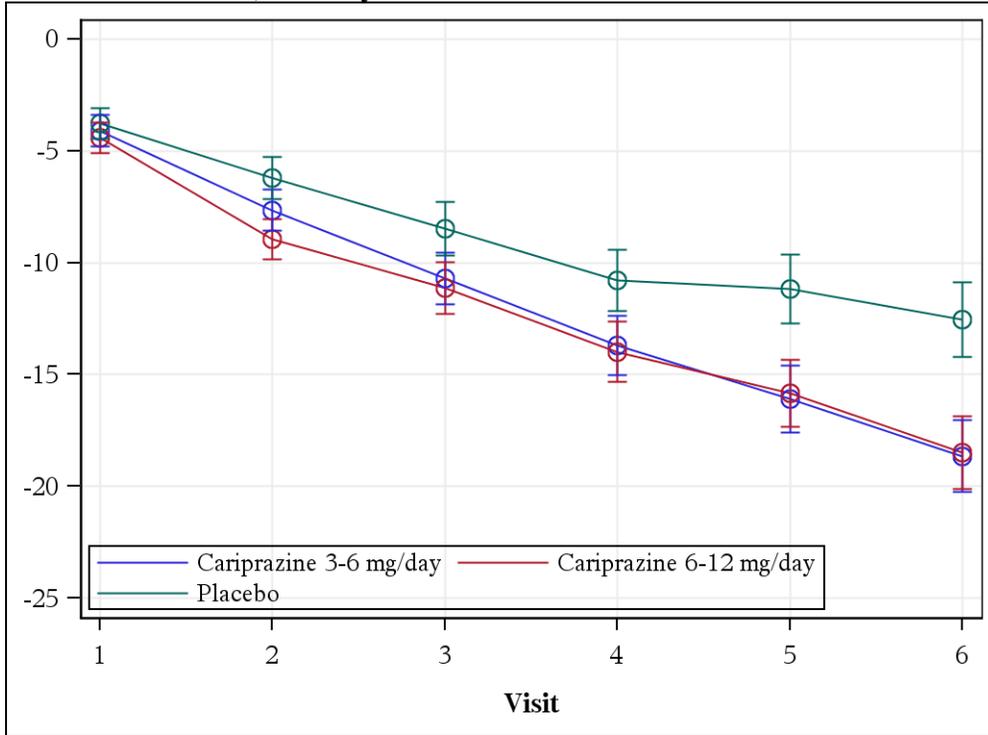
Figure 12: Visit-wise LS mean estimates and 95% Confidence Intervals (Change from Baseline in YMRS) – Study MD-32



[Source: Reviewer's results]

Note: Visit 5 is the 3-week endpoint.

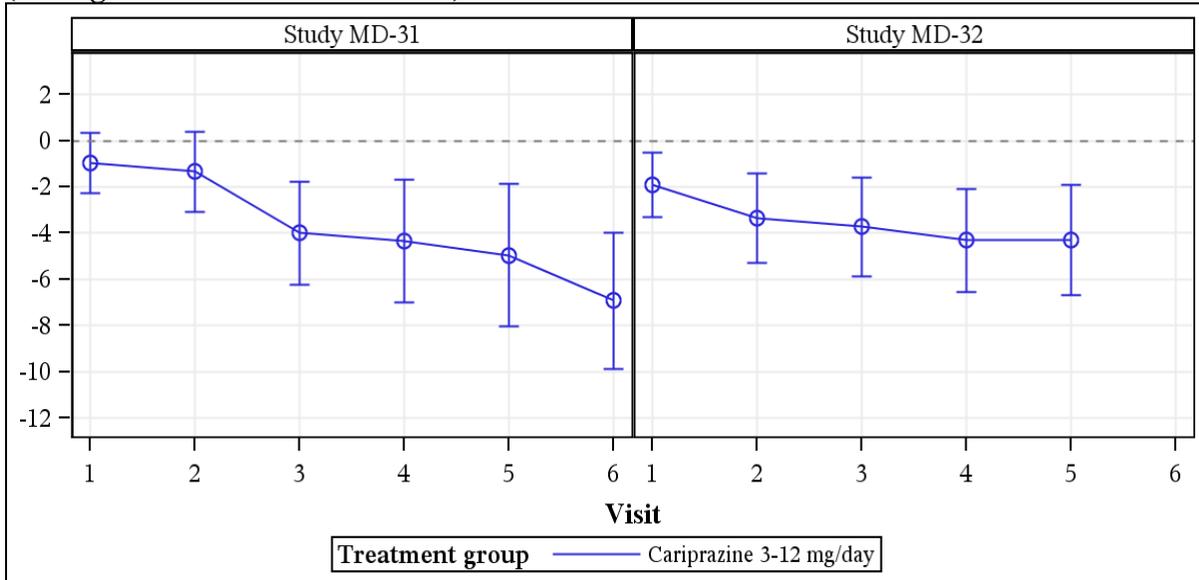
Figure 13: Visit-wise LS mean estimates and 95% Confidence Intervals (Change from Baseline in YMRS) – Study MD-33



[Source: Reviewer's results]

Note: Visit 6 is the 3-week endpoint.

Figure 14: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in YMRS) – Studies MD-31 and MD-32



[Source: Reviewer's results]

Note: Visit 6 is the 3-week endpoint.

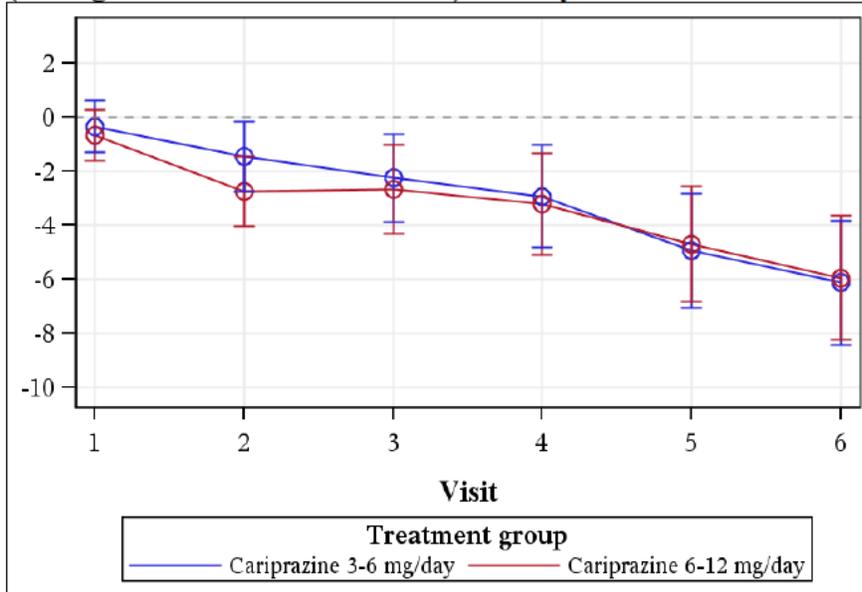
Table 14: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in YMRS) – Bipolar Mania pivotal studies

Study	Treatment arm	Visit	LS Mean Difference	Standard Error	95% Confidence Interval
RGH-MD-31	Cariprazine 3-12 mg/day	1	-0.9	0.7	(-2.3, 0.4)
		2	-1.3	0.9	(-3.1, 0.4)
		3	-4.0	1.1	(-6.2, -1.8)
		4	-4.3	1.3	(-7.0, -1.7)
		5	-5.0	1.6	(-8.1, -1.9)
		6	-6.9	1.5	(-9.9, -4.0)
RGH-MD-32	Cariprazine 3-12 mg/day	1	-1.9	0.7	(-3.3, -0.5)
		2	-3.3	1.0	(-5.3, -1.4)
		3	-3.7	1.1	(-5.9, -1.6)
		4	-4.3	1.1	(-6.5, -2.1)
		5	-4.3	1.2	(-6.7, -1.9)
RGH-MD-33	Cariprazine 3-6 mg/day	1	-0.3	0.5	(-1.3, 0.6)
		2	-1.4	0.7	(-2.7, -0.1)
		3	-2.2	0.8	(-3.9, -0.6)
		4	-2.9	1.0	(-4.8, -1.0)
		5	-4.9	1.1	(-7.1, -2.8)
		6	-6.1	1.2	(-8.4, -3.8)
	Cariprazine 6-12 mg/day	1	-0.7	0.5	(-1.6, 0.3)
		2	-2.7	0.7	(-4.0, -1.4)
		3	-2.7	0.8	(-4.3, -1.0)
		4	-3.2	1.0	(-5.1, -1.3)
		5	-4.7	1.1	(-6.8, -2.5)
		6	-5.9	1.2	(-8.2, -3.6)

[Source: Reviewer's results]

Note: Visit 5 of Study MD-32 and Visit 6 of other studies were Week 3 endpoint.

Figure 15: Visit-wise LS mean differences from Placebo and 95% Confidence Intervals (Change from Baseline in YMRS) – Study MD-33



[Source: Reviewer's results]

Note: Visit 6 is the 3-week endpoint.

Table 15: MMRM and ANCOVA (LOCF) results for key secondary efficacy (CGI-S)

[Source: Sponsor's Clinical Study Reports and Reviewer's results]

Study (# subjects of ITT population)	Treatment arms (# subjects of ITT population)	Mean Baseline score (SD)	MMRM		ANCOVA (LOCF)	
			LS Mean of change from baseline (SE)	LS Mean Difference from Placebo	LS Mean of change from baseline (SE)	LS Mean Difference from Placebo
MD-31 (N=235)	Cariprazine 3-12 mg/day (N=118)	4.7 (1.1)	-1.6 (0.1)	-0.7	-1.6 (0.1)	-0.7
	Placebo (N=117)	4.6 (1.1)	-0.9 (0.1)	-	-0.9 (0.1)	-
MD-32 (N=310)	Cariprazine 3-12 mg/day (N=158)	4.6 (0.6)	-1.9 (0.1)	-0.4	-1.6 (0.1)	-0.4
	Placebo (N=152)	4.6 (0.6)	-1.5 (0.1)	-	-1.3 (0.1)	-
MD-33 (N=492)	Cariprazine 3-6 mg/day (N=165)	4.8 (0.6)	-1.9 (0.1)	-0.6	-1.8 (0.1)	-0.6
	Cariprazine 6-12 mg/day (N=167)	4.8 (0.6)	-1.9 (0.1)	-0.6	-1.7 (0.1)	-0.5
	Placebo (N=160)	4.8 (0.7)	-1.3 (0.1)	-	-1.2 (0.1)	-

Note: (1) The bolded results are from pre-specified primary analysis. (2) In Study RGH-MD-31 did not include an interaction term of Baseline score and Visit. This reviewer confirmed the results are almost identical when this interaction term is included. (3) Adjusted *p* values for the key secondary endpoints are given in Table 11.

3.2.4.3 Center impacts on efficacy

In this section, the impacts of study centers on primary efficacy are examined. For each study (of both programs of schizophrenia and bipolar I disorder), by treatment group within every center, placebo-subtracted mean change from baseline scores to endpoint were calculated based on primary efficacy raw data. The calculated means are unadjusted and simple averages of these primary efficacy variables. For each indication, two figures are presented below. One figure shows that for each study, the country-wise plots of the placebo-subtracted mean change scores of treatment groups within each center. This figure helps us see variations of efficacy assessments by study centers (treatment groups of every center when there are more than one treatment group exists) for each country within study. Another figure presents the plots of the placebo-subtracted mean change scores against the sample sizes of centers (treatment groups of every center). It helps us see if any study center with a large sample size impacted the overall efficacy assessment.

From a visual inspection of Figure 16 and Figure 18, in both indications, this reviewer found that across the studies, center-wise efficacy assessments were fairly well-balanced among countries. In short, no particular country had an unusual efficacy assessment results that may affect the overall efficacy assessment.

From a visual inspection of Figure 17 and Figure 19, there were only a few centers with a large sample size that had a very positive (drug favoring) value of placebo-subtracted mean change from baseline score. Consequently, without these centers, the primary efficacy results were found to still hold.

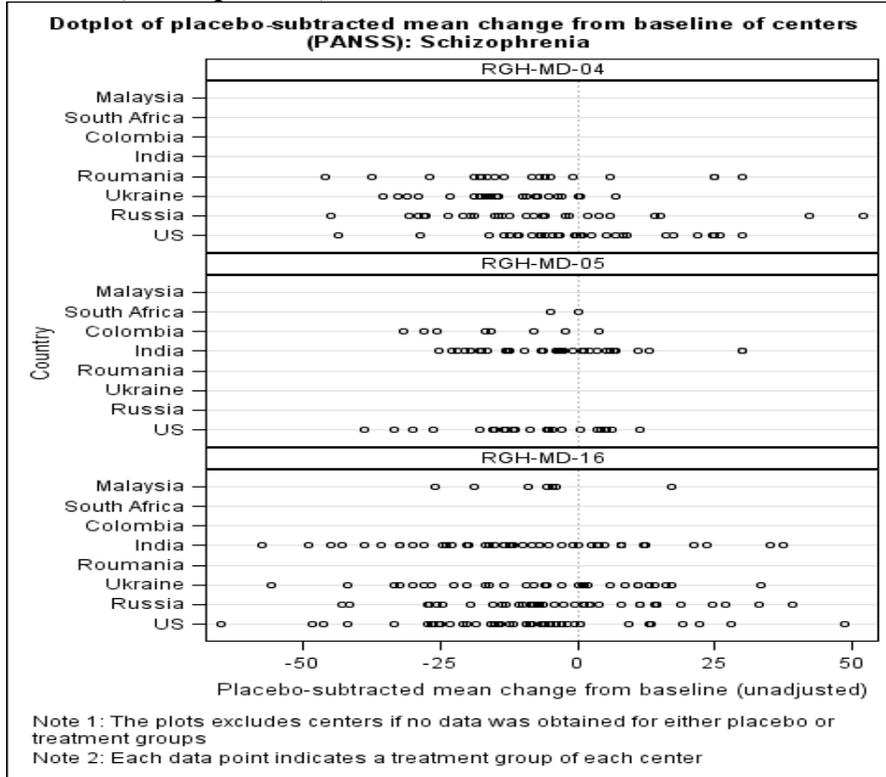
3.2.4.3.1 *Schizophrenia*

As shown in Figure 16, the distributions of center-wise efficacy assessment result on PANSS were fairly comparable among countries. As seen in Figure 17, only a very few centers with a large sample size of Study MD-05 had PANSS efficacy assessment results with a relatively large deviation from zero.

3.2.4.3.2 *Bipolar I disorder*

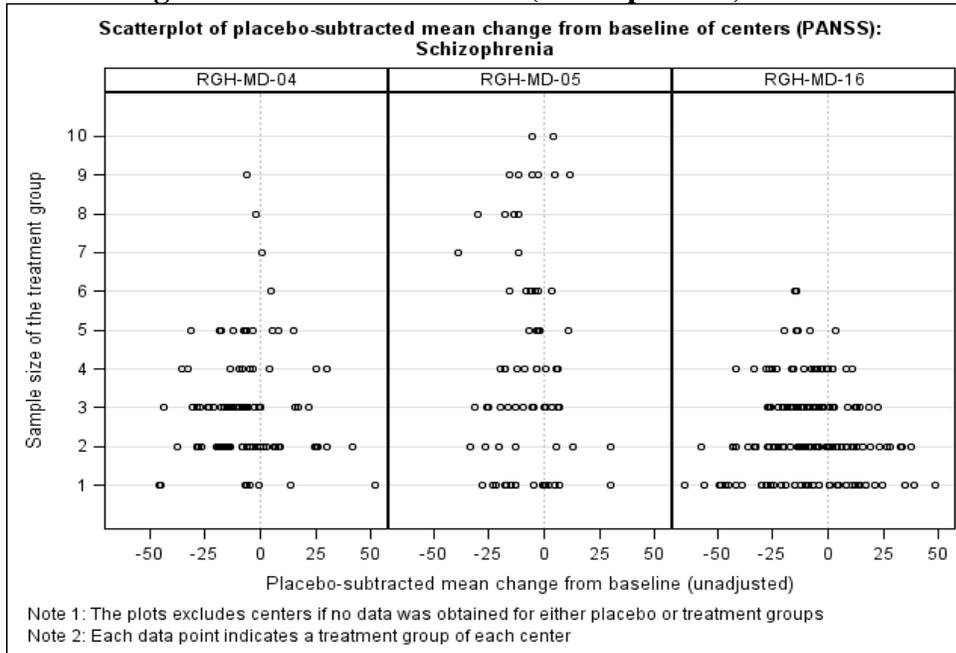
As shown in Figure 18, the distributions of center-wise efficacy assessment result on PANSS were fairly comparable among countries, except for US study centers. In all three studies, US study centers had a much narrower range of placebo-subtracted change from baseline scores in YMRS assessments. The LS mean difference in change from baseline of US was relatively lower in four treatment arms of the three studies than that of any other country. (This excludes Croatia/Serbia, since they had very few subjects). See also subgroup analysis for Country (Table 17). As seen in Figure 19, in all three studies, only a very few centers with a large sample size had YMRS efficacy assessment results with a relatively large deviation from zero.

Figure 16: Center Impact by Country: Placebo-subtracted mean change from baseline in PANSS (Schizophrenia)



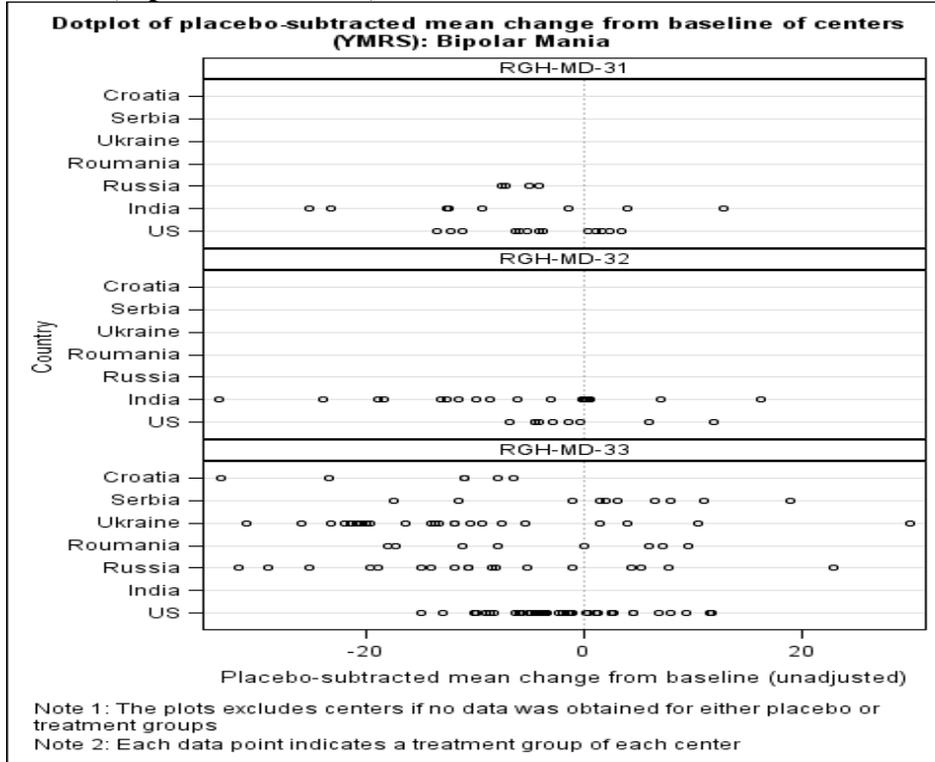
[Source: Reviewer's results]

Figure 17: Center Impact by Size of Treatment groups within Center: Placebo-subtracted mean change from baseline in PANSS (Schizophrenia)



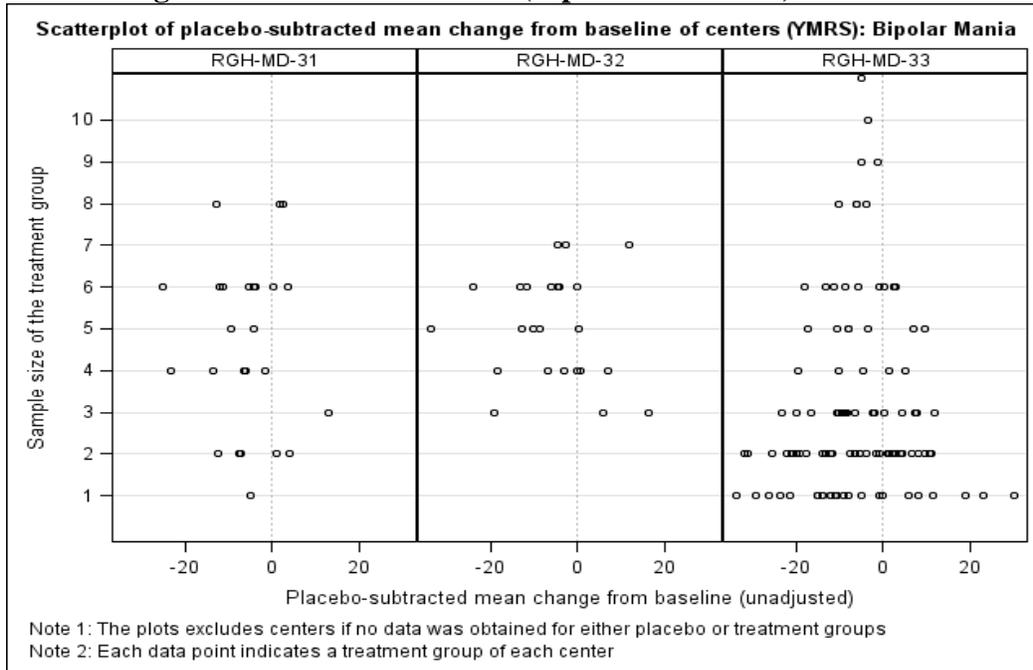
[Source: Reviewer's results]

Figure 18: Center Impact by Country: Placebo-subtracted mean change from baseline in YMRS (Bipolar I Disorder)



[Source: Reviewer's results]

Figure 19: Center Impact by Size of Treatment groups within Center: Placebo-subtracted mean change from baseline in YMRS (Bipolar I Disorder)



[Source: Reviewer's results]

3.3 Evaluation of Safety

Safety was not evaluated in this review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age group, and Country

In this section, efficacy results of three schizophrenia and three bipolar pivotal studies across subgroups (gender, race, age group and country) are presented.

For each study, the estimated LS mean difference from placebo in change from baseline is plotted and tabulated for each level of subgroup and for the entire treatment group. The estimates are based on MMRM with treatment, study center, visit, treatment-by-visit interaction as factors, baseline score and baseline score-by-visit interaction as covariates. The 95% confidence intervals are given only for the overall treatment arm. The levels of age group (< 30 years old, ≥ 30 and <50 years old, ≥ 50 years old) are not based on clinical considerations. The tables of these estimates include the numbers of subjects of treatment and placebo arms by subgroup.

Differences in efficacy due to any of the subgroups based on gender, race and age are not presumed in the submitted NDA studies, as in most clinical studies. The observed differences of LS mean estimates of subgroups should not be used to conclude differences or sameness in efficacy among subgroups. As can be seen in the following tables, a few of the LS mean estimates of subgroups may not be even within the 95% confidence interval of the overall population. These differences may be interpreted as being due to the fact that the sample sizes of subgroups were small, and/or due to pure randomness. In both indications, we did not see any tendency that is consistent enough in efficacy estimates for subgroup of *gender*, *race* or *age group* for us to suspect efficacy differences in these subgroups.

It is noted that in general, countries may show some differences in efficacy due to real, and clinically meaningful, differences. One may believe that *country* is important in this sense in a particular study. If this was the case, a country's efficacy estimates with a fairly large sample size might be expected to exhibit some consistency across multiple studies (say, one country is consistently better in efficacy estimates than any other countries). Then the observed differences may be taken as suggesting clinically relevant, if not statistically definitive, information on efficacy. In the three cariprazine bipolar studies, the United States exhibited relatively lower efficacy in LS mean estimates. This observation should not directly lead to a conclusion that cariprazine is less efficacious in US bipolar mania patients than in the other countries' patients who participated in the cariprazine bipolar studies, but may be worth mentioning.

4.1.1 Schizophrenia Pivotal Studies

4.1.1.1 Age group

The sample size of the age group 50 years old or older was relatively smaller in Studies MD-05 and MD-16. The observed efficacy estimates of the less than 30 years old age range were

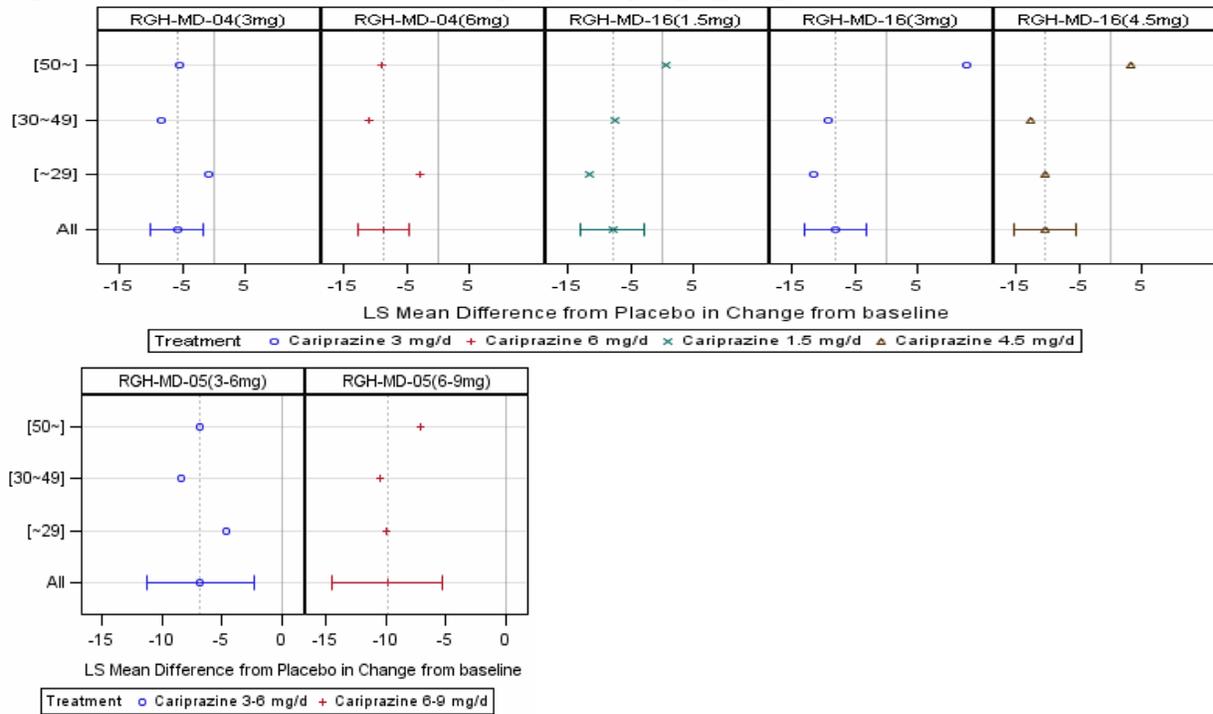
relatively lower in Study MD-04, but relatively higher in Study MD-16. In Study MD-05 (a flexible dosing study), the efficacy estimates of all age ranges were good perhaps due to optimized dosing (Table 16 and Figure 20).

Table 16: LS mean difference from placebo by Age group

Study	Treatment arm	Age group	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline	
RGH-MD-04	Cariprazine 3 mg/day	All	151	149	-6.0 (-10.1,-1.9)	
		[-29]	38	44	-0.9	
		[30~49]	88	73	-8.6	
		[50~]	25	32	-5.5	
	Cariprazine 6 mg/day	All	154	149	-8.8 (-12.9,-4.7)	
		[-29]	40	44	-2.9	
		[30~49]	88	73	-11.1	
		[50~]	26	32	-8.9	
RGH-MD-16	Cariprazine 1.5 mg/day	All	140	148	-8.0 (-12.9,-3.0)	
		[-29]	47	59	-11.7	
		[30~49]	82	75	-7.7	
		[50~]	11	14	0.5	
	Cariprazine 3 mg/day	All	140	148	-8.2 (-13.1,-3.2)	
		[-29]	45	59	-11.6	
		[30~49]	82	75	-9.3	
		[50~]	13	14	12.6	
	Cariprazine 4.5 mg/day	All	145	148	-10.5 (-15.4,-5.6)	
		[-29]	54	59	-10.6	
		[30~49]	76	75	-12.9	
		[50~]	15	14	3.0	
	RGH-MD-05	Cariprazine 3-6 mg/day	All	147	145	-6.8 (-11.3,-2.4)
			[-29]	40	52	-4.7
			[30~49]	85	73	-8.4
			[50~]	22	20	-6.9
Cariprazine 6-9 mg/day		All	147	145	-9.9 (-14.5,-5.3)	
		[-29]	47	52	-9.9	
		[30~49]	87	73	-10.4	
		[50~]	13	20	-7.1	

[Source: Reviewer's results]

Figure 20: LS mean difference from placebo by Age group



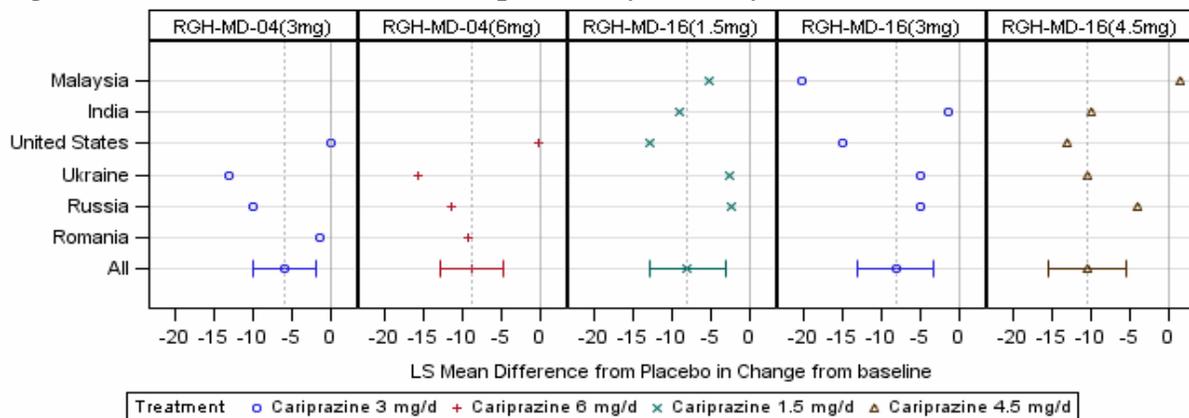
[Source: Reviewer’s results]

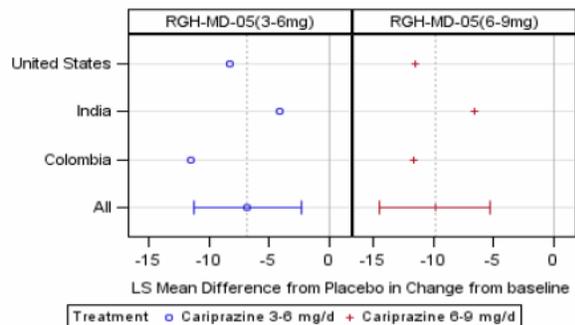
Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 16. The age groups are labeled as [~29], [30~49], [50~] for less than 30 years old, 30 years old or older and less than 50 years old, 50 years old or older, respectively.

4.1.1.2 Country

In Studies MD-16 and MD-05, US subjects performed much better than subjects of any other country (The sample sizes of Colombia and Malaysia were small). But in Study MD-04, US subjects performed poorly. We do not see any consistent differences in efficacy results among countries (See Table 17 and Figure 21).

Figure 21: LS mean difference from placebo by Country





[Source: Reviewer's results]

Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 17.

Table 17: LS mean difference from placebo by Country

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-04	Cariprazine 3 mg/day	All	151	149	-6.0 (-10.1,-1.9)
		Romania	18	14	-1.3
		Russia	46	47	-10.0
		Ukraine	34	36	-13.1
		United States	53	52	0.2
	Cariprazine 6 mg/day	All	154	149	-8.8 (-12.9,-4.7)
		Romania	19	14	-9.3
		Russia	45	47	-11.5
		Ukraine	37	36	-15.7
		United States	53	52	-0.2
RGH-MD-16	Cariprazine 1.5 mg/day	All	140	148	-8.0 (-12.9,-3.0)
		India	30	29	-9.0
		Malaysia	3	4	-5.2
		Russia	32	35	-2.4
		Ukraine	24	23	-2.5
		United States	51	57	-12.9
	Cariprazine 3 mg/day	All	140	148	-8.2 (-13.1,-3.2)
		India	31	29	-1.4
		Malaysia	4	4	-20.2
		Russia	31	35	-4.9
		Ukraine	22	23	-5.0
United States		52	57	-14.9	

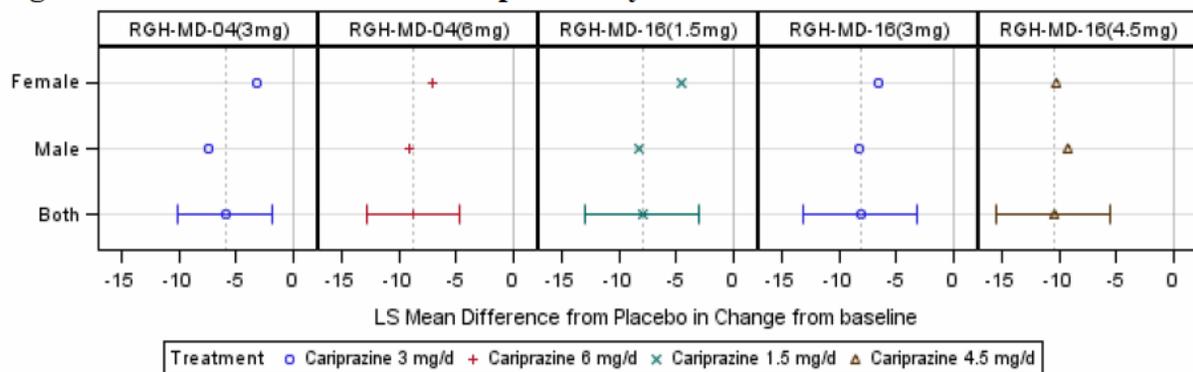
Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
	Cariprazine 4.5 mg/day	All	145	148	-10.5 (-15.4,-5.6)
		India	34	29	-9.9
		Malaysia	5	4	1.5
		Russia	31	35	-4.0
		Ukraine	23	23	-10.4
		United States	52	57	-13.1
RGH-MD-05	Cariprazine 3-6 mg/day	All	147	145	-6.8 (-11.3,-2.4)
		Colombia	16	16	-11.6
		India	53	54	-4.2
		United States	76	74	-8.2
	Cariprazine 6-9 mg/day	All	147	145	-9.9 (-14.5,-5.3)
		Colombia	15	16	-11.7
		India	54	54	-6.6
		United States	77	74	-11.5

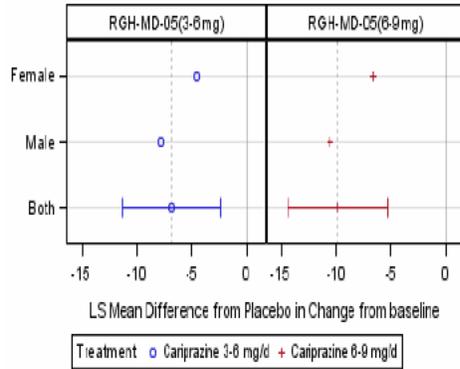
[Source: Reviewer's results]

4.1.1.3 Gender

There are numerical differences by gender observed in efficacy estimates; males appear to have performed better than females in almost all treatment arms of the three studies (See Figure 22 and Table 18).

Figure 22: LS mean difference from placebo by Gender





[Source: Reviewer's results]

Note: The plots are LS mean difference from placebo in change from baseline based on MMRM (Table 18).

Table 18: LS mean difference from placebo by Gender

Study	Treatment arm	Gender	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-04	Cariprazine 3 mg/day	Both	151	149	-6.0 (-10.1,-1.9)
		Male	95	93	-7.3
		Female	56	56	-3.2
	Cariprazine 6 mg/day	Both	154	149	-8.8 (-12.9,-4.7)
		Male	97	93	-9.1
		Female	57	56	-7.1
RGH-MD-16	Cariprazine 1.5 mg/day	Both	140	148	-8.0 (-12.9,-3.0)
		Male	91	100	-8.2
		Female	49	48	-4.5
	Cariprazine 3 mg/day	Both	140	148	-8.2 (-13.1,-3.2)
		Male	101	100	-8.3
		Female	39	48	-6.6
	Cariprazine 4.5 mg/day	Both	145	148	-10.5 (-15.4,-5.6)
		Male	102	100	-9.3
		Female	43	48	-10.3
RGH-MD-05	Cariprazine 3-6 mg/day	Both	147	145	-6.8 (-11.3,-2.4)
		Male	115	108	-7.8
		Female	32	37	-4.5
	Cariprazine 6-9 mg/day	Both	147	145	-9.9 (-14.5,-5.3)
		Male	112	108	-10.6
		Female	35	37	-6.6

[Source: Reviewer's results]

4.1.1.4 Race

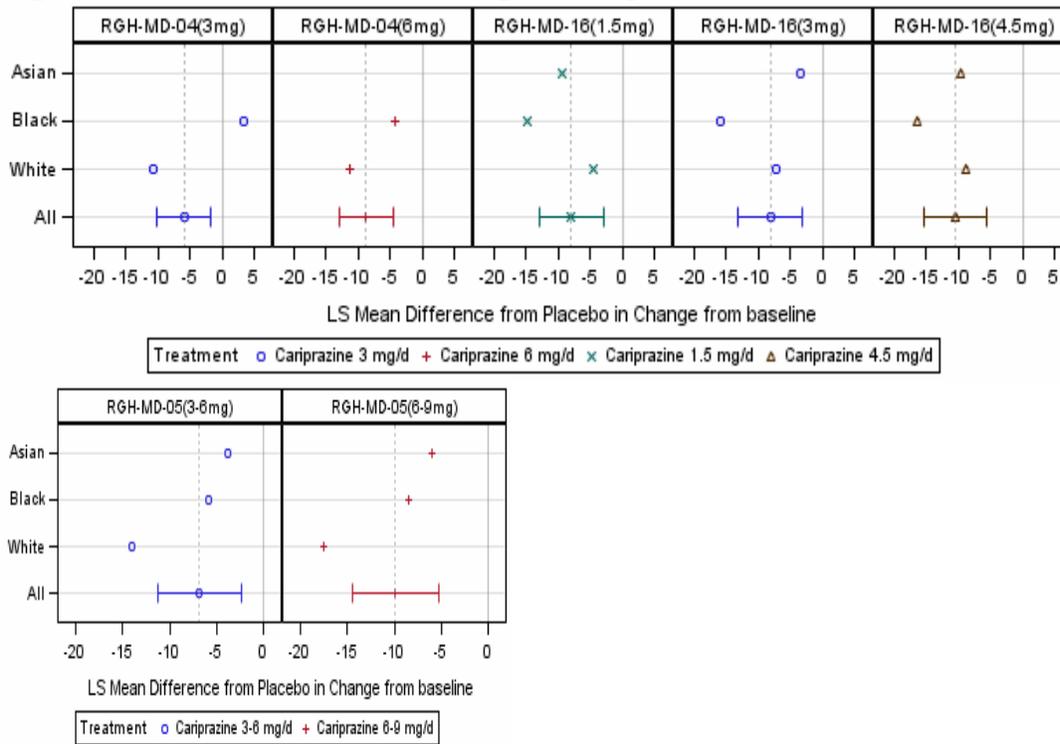
The efficacy estimates of White subjects were better in Study MD-04 than those of Black subjects, and better in Study MD-05 than those of Black and Asian subjects. On the other hand, those of Black subjects were better in Study MD-16 than those of White and Asian subjects.

Table 19: LS mean difference from placebo by Race

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-04	Cariprazine 3 mg/day	All	151	149	-6.0 (-10.1,-1.9)
		White	100	93	-10.8
		Black	31	39	3.4
	Cariprazine 6 mg/day	All	154	149	-8.8 (-12.9,-4.7)
		White	101	93	-11.2
		Black	34	39	-4.4
RGH-MD-16	Cariprazine 1.5 mg/day	All	140	148	-8.0 (-12.9,-3.0)
		White	74	80	-4.6
		Black	31	34	-14.7
		Asian	34	33	-9.3
	Cariprazine 3 mg/day	All	140	148	-8.2 (-13.1,-3.2)
		White	70	80	-7.2
		Black	35	34	-15.9
		Asian	35	33	-3.5
	Cariprazine 4.5 mg/day	All	145	148	-10.5 (-15.4,-5.6)
		White	73	80	-8.8
		Black	32	34	-16.3
		Asian	39	33	-9.6
RGH-MD-05	Cariprazine 3-6 mg/day	All	147	145	-6.8 (-11.3,-2.4)
		White	28	25	-14.0
		Black	53	51	-5.9
		Asian	55	55	-3.8
	Cariprazine 6-9 mg/day	All	147	145	-9.9 (-14.5,-5.3)
		White	30	25	-17.5
		Black	53	51	-8.5
		Asian	55	55	-6.1

[Source: Reviewer's results]

Figure 23: LS mean difference from placebo by Race



[Source: Reviewer's results]

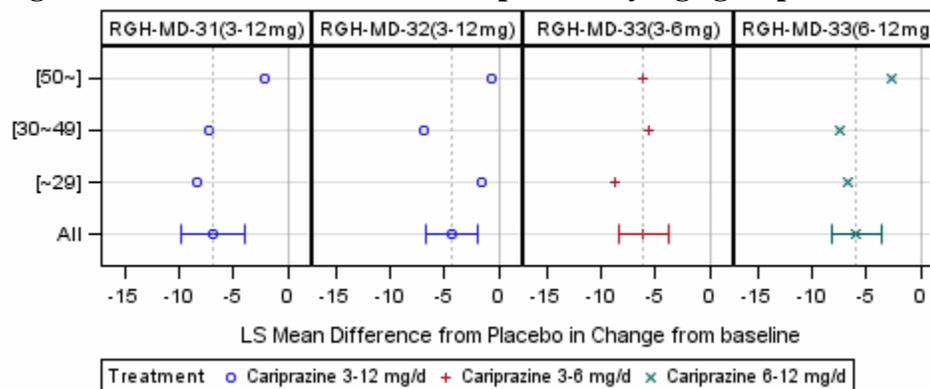
Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 19. Due to their small sample size, the results of 'Other' (other races) are not included.

4.1.2 Bipolar I Disorder Pivotal Studies

4.1.2.1 Age group

In all three studies, the efficacy estimates of subjects of 50 years old or older exhibited relatively lower efficacy, especially when the sample size was small, compared to those of younger age.

Figure 24: LS mean difference from placebo by Age group



[Source: Reviewer's results]

Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 20. The age groups are labeled as [~29], [30~49], [50~] for less than 30 years old, 30 years old or older and less than 50 years old, 50 years old or older, respectively.

Table 20: LS mean difference from placebo by Age group

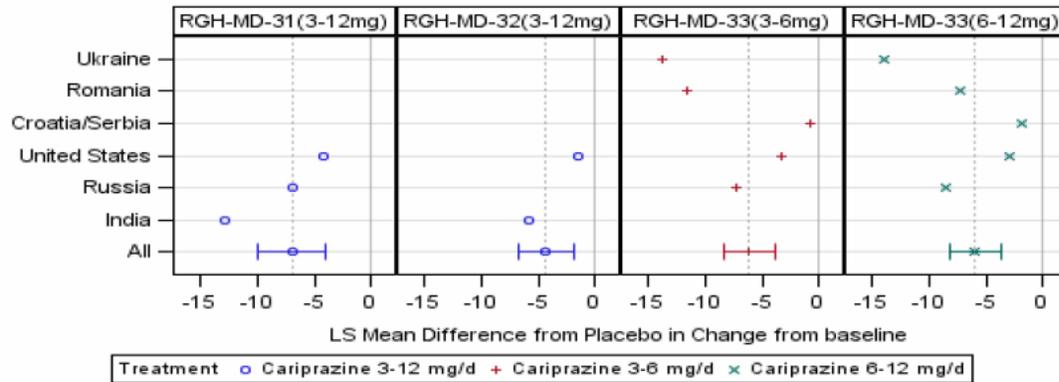
Study	Treatment arm	Age group	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-31	Cariprazine 3-12 mg/day	All	118	117	-6.9 (-9.9,-4.0)
		[~29]	31	30	-8.5
		[30~49]	72	67	-7.2
		[50~]	15	20	-2.1
RGH-MD-32	Cariprazine 3-12 mg/day	All	158	152	-4.3 (-6.7,-1.9)
		[~29]	53	51	-1.6
		[30~49]	83	75	-6.9
		[50~]	22	26	-0.7
RGH-MD-33	Cariprazine 3-6 mg/day	All	165	160	-6.1 (-8.4,-3.8)
		[~29]	31	33	-8.7
		[30~49]	77	81	-5.6
		[50~]	57	46	-6.2
	Cariprazine 6-12 mg/day	All	167	160	-5.9 (-8.2,-3.6)
		[~29]	29	33	-6.8
		[30~49]	91	81	-7.4
		[50~]	47	46	-2.7

[Source: Reviewer's results]

4.1.2.2 Country

The efficacy estimates of US subjects exhibited lower efficacy than subjects of any other country. In Study MD-31, this reviewer repeated the primary analysis without India, and in MD-33 without Ukraine. The results suggested that the statistical conclusions drawn for the primary analysis would not be changed without India (in Study MD-31) and without Ukraine (in Study MD-33). Due to their small sample size, data of Croatia and Serbia were merged.

Figure 25: LS mean difference from placebo by Country



[Source: Reviewer's results]

Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 21.

Table 21: LS mean difference from placebo by Country

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-31	Cariprazine 3-12 mg/day	All	118	117	-6.9 (-9.9,-4.0)
		India	34	32	-12.7
		Russia	11	12	-7.0
		United States	73	73	-4.3
RGH-MD-32	Cariprazine 3-12 mg/day	All	158	152	-4.3 (-6.7,-1.9)
		India	91	87	-5.9
		United States	67	65	-1.5
RGH-MD-33	Cariprazine 3-6 mg/day	All	165	160	-6.1 (-8.4,-3.8)
		Croatia/Serbia	11	11	-0.8
		Romania	17	15	-11.5
		Russia	26	26	-7.3
		Ukraine	22	21	-13.7
		United States	89	87	-3.3

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
	Cariprazine 6-12 mg/day	All	167	160	-5.9 (-8.2,-3.6)
		Croatia/Serbia	14	11	-1.8
		Romania	17	15	-7.3
		Russia	26	26	-8.6
		Ukraine	23	21	-14.0
		United States	87	87	-2.9

[Source: Reviewer's results]

4.1.2.3 Gender

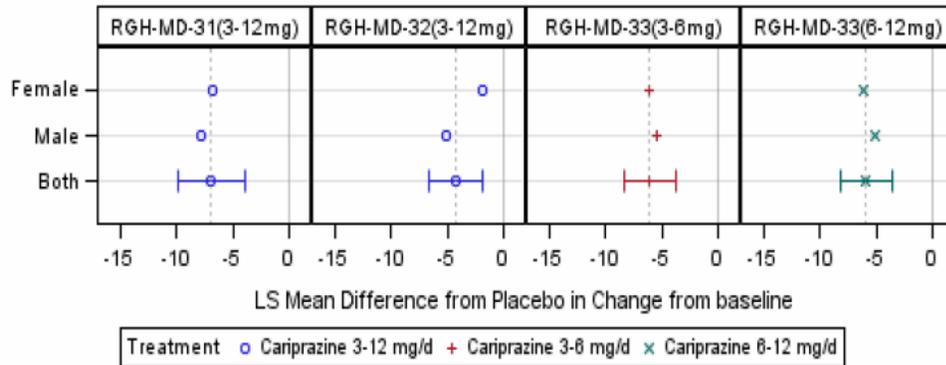
The efficacy estimates of the three studies showed no difference between males and females.

Table 22: LS mean difference from placebo by Gender

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-31	Cariprazine 3-12 mg/day	Both	118	117	-6.9 (-9.9,-4.0)
		Male	80	76	-7.8
		Female	38	41	-6.8
RGH-MD-32	Cariprazine 3-12 mg/day	Both	158	152	-4.3 (-6.7,-1.9)
		Male	105	94	-5.1
		Female	53	58	-1.9
RGH-MD-33	Cariprazine 3-6 mg/day	Both	165	160	-6.1 (-8.4,-3.8)
		Male	89	89	-5.4
		Female	76	71	-6.2
	Cariprazine 6-12 mg/day	Both	167	160	-5.9 (-8.2,-3.6)
		Male	84	89	-5.1
		Female	83	71	-6.1

[Source: Reviewer's results]

Figure 26: LS mean difference from placebo by Gender



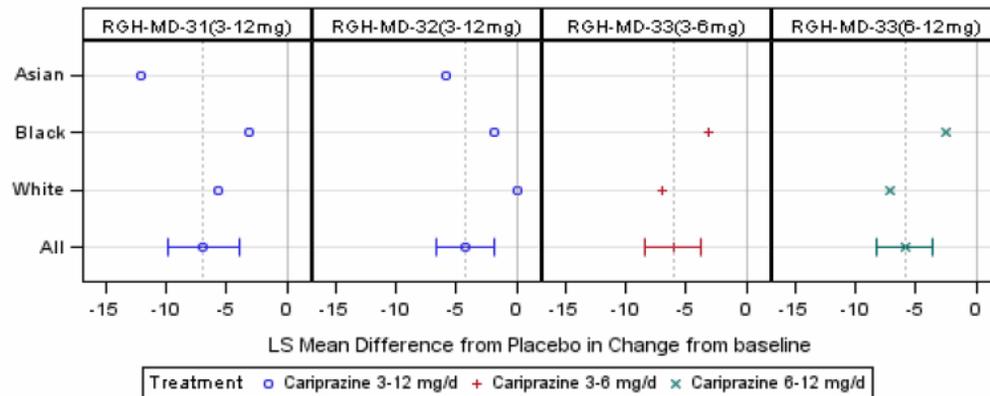
[Source: Reviewer’s results]

Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 22.

4.1.2.4 Race

No consistent differences in the efficacy estimates from the three studies were observed between race subgroups.

Figure 27: LS mean difference from placebo by Race



[Source: Reviewer’s results]

Note: The plots are based on LS mean difference from placebo in change from baseline based on MMRM, which are listed in Table 23. Due to their small sample size, the results of ‘Other’ (other races) are not included.

Table 23: LS mean difference from placebo by Race

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-31	Cariprazine 3-12 mg/day	All	118	117	-6.9 (-9.9,-4.0)
		White	47	55	-5.7
		Black	36	30	-3.1
		Asian	30	28	-12.1

Study	Treatment arm	Race	#subjects (Treatment arm)	#subjects (Placebo arm)	LS Mean Difference from Placebo in Change from baseline
RGH-MD-32	Cariprazine 3-12 mg/day	All	158	152	-4.3 (-6.7,-1.9)
		White	33	33	0.0
		Black	33	28	-1.8
		Asian	91	87	-5.9
RGH-MD-33	Cariprazine 3-6 mg/day	All	165	160	-6.1 (-8.4,-3.8)
		White	117	113	-7.0
		Black	43	44	-3.1
	Cariprazine 6-12 mg/day	All	167	160	-5.9 (-8.2,-3.6)
		White	112	113	-7.1
		Black	51	44	-2.5

[Source: Reviewer's results]

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

5.1.1 Schizophrenia

In the schizophrenia program, three pivotal studies have shown efficacy of cariprazine:

- In two fixed dose studies, efficacy of daily fixed doses of cariprazine 1.5 mg, 3 mg, 4.5 mg and 6 mg was established at the 6-week endpoint based on PANSS total score and CGI-S score.
- In one fixed-flexible dose study, efficacy of cariprazine dose ranges of 3-6 mg and 6-9 mg was established at the 6-week endpoint based on PANSS total score and CGI-S score.

One supportive study did not achieve statistical significance. However, it has shown a trend for efficacy of cariprazine of a dose range of 3-12 mg at the 6-week endpoint based on PANSS total score and CGI-S score.

It is noted that efficacy of a daily fixed dose of 9 mg has not been fully evaluated in the clinical program, although there were some patients in the phase-3 fixed-flexible dose study who reached the upper-end dose (9 mg) of a dose range of 6-9 mg for their optimized dose. The efficacy estimates from a fixed-flexible dose study may not serve as a basis for establishing efficacy of fixed doses. In dose optimization as conducted in the fixed-flexible dose studies, it may not be feasible to evaluate “efficacy difference due to different doses”. By design, only part of randomized patients assigned to the 6-9 mg dose range reached 9 mg for their optimal dose (Table 7). In general, observed efficacy of an optimal dose may not be sufficient for understanding efficacy of a fixed dose. Furthermore, the dose range was not designed to be compared with any fixed dose within the same study. This reviewer notes that the FDA

communicated to the sponsor its preference for studies looking at fixed doses within the proposed fixed-flexible dose study, at Type C Guidance meeting held on February 11, 2009¹⁸. This reviewer also points out that, at this meeting, the sponsor asked the FDA if an approval might be possible for a dose range. The FDA responded by stating that it would be possible, but as mentioned above, cautioned that efficacy information of fixed doses would be important in efficacy assessments. Beyond statistical considerations, clinically it may be meaningful to consider an approval for a dose-range. This reviewer notes that the proportion of subjects who reached the upper-end dose (of a dose-range), their efficacy outcome and overall mean daily exposure were not consistent between the two fixed-flexible dose studies (Table 7). See *Section 3.2.4.1.2 Reviewer's assessments* for more details.

The efficacy estimates of fixed doses of 1.5 mg, 3 mg, 4.5 mg and 6 mg do not seem to suggest a dose-response relationship across the two fixed dose studies. Dose response may have been observed in each of the two fixed dose studies, but no consistent dose responses for the fixed doses were observed across these studies (Figure 1 and Figure 2).

In the fixed-flexible dose study (MD-05), the efficacy estimate of 6-9 mg was better than that of 3-6 mg (Table 8). However, an efficacy comparison of dose ranges may not be meaningful when dose was optimized.

5.1.2 Bipolar I Disorder

This reviewer notes that no information on efficacy of fixed doses is available in the submitted NDA study data. At Type C meeting (see footnote 1), the sponsor asked whether it would be sufficient to demonstrate efficacy in two flexible dose studies in mania, as opposed to one fixed-dose study and one flexible-dose study. The FDA cautioned that assessing effective doses without fixed dose studies may become difficult, but agreed that it would be acceptable for the sponsor to conduct one flexible dose study in mania and one fixed-flexible dose study in mania.

A dose response relationship of fixed doses may be difficult to assess from the submitted efficacy data, since there is no information on efficacy of fixed doses in the bipolar studies. See *Section 3.2.4.2.2 Reviewer's assessments* for more details.

A differential in efficacy estimates between the two dose-ranges, 3-6 mg and 6-12 mg, may be viewed as negligible. The efficacy estimate of the 3-12 mg dose range was much better in MD-32 than in MD-31, but the proportion of subjects of 3-12 mg dose-range who reached the target (upper-end) dose of 12 mg was much higher in MD-31 than in MD-32 (See Table 12 and Table 13). Furthermore, dose increase decisions were made differently in the same dose range (3-12 mg) between the two studies MD-31 and MD-32 (See *Section 3.2.4.2.2 Reviewer's assessments*). The two dose-ranges of Study MD-33 (3-6 mg and 6-12 mg) exhibited almost no difference in efficacy estimates. This may be natural as dose was optimized. It may not be meaningful to draw a conclusion on dose response of the dose-ranges.

¹⁸ Preliminary Comments/Minutes of Meeting IND (b)(4) [Schizophrenia], IND 77726 [bipolar]: RGH-188 (cariprazine). The meeting minutes document is available in DARRTS. It was finalized on February 20, 2009.

Note: In both indications, the above efficacy conclusions are based on statistical evidence obtained from efficacy studies whose study designs and analysis plans were agreed upon between the sponsor and the FDA. The efficacy estimates of the primary and key secondary endpoints are displayed in Table 8 and Table 10 for schizophrenia and in Table 13 and Table 15 for bipolar mania.

5.2 Conclusions and Recommendations

5.2.1 Schizophrenia

The cariprazine schizophrenia NDA program of Forest includes three pivotal studies and one supportive study for efficacy assessment of the candidate drug. Two pivotal fixed dose studies of this program have statistically established evidence that daily fixed doses of cariprazine 1.5 mg, 3 mg, 4.5 mg and 6 mg are efficacious. One pivotal fixed-flexible dose study of the program has established statistical evidence that two cariprazine dose ranges of 3-6 mg and 6-9 mg were efficacious in optimized dose within the fixed dose range. Another proof-of-concept, fixed-flexible dose study with two dose-ranges 1.5-4.5 mg and 6-12 mg, submitted as a supportive study, has shown a trend for cariprazine efficacy, although it failed in showing statistical significance.

The sponsor considers doses of 1.5 to (b) (4) mg administered once daily effective. (b) (4)

Based on statistical evidence, this reviewer suggests that the doses of 1.5 mg, 3 mg, 4.5 mg and 6 mg may be included in the product label as *effective doses*, as an effective dose is usually meant for efficacy of a fixed daily dose in the schizophrenia product label. A dose response relationship has not been consistently observed across the two fixed dose studies.

Given the positive results of the fixed-flexible dose study, an approval of the two dose-ranges as *effective dose ranges* may also be considered. The appropriateness of such a decision may be a clinical matter. It may not be appropriate to draw any conclusion on a dose response relationship, based on the efficacy estimates of the dose ranges (3-6 mg and 6-9 mg).

5.2.2 Bipolar I Disorder

The cariprazine bipolar I disorder NDA program of Forest includes three pivotal studies, i.e., two flexible dose studies and one fixed-flexible dose study for efficacy assessment of the candidate drug. A dose-range of cariprazine 3-12 mg has been statistically shown efficacious in the two pivotal flexible dose studies. Furthermore, two dose-ranges of cariprazine 3-6 mg and 6-12 mg have been statistically shown efficacious in the pivotal fixed-flexible dose study.

It may be a clinical matter how efficacy of the dose-ranges should be described in the label. As no efficacy data of fixed doses in bipolar mania are available, the submitted data do not admit a statistical assessment of efficacy of fixed doses.

This reviewer cautions that it may not be appropriate to draw any conclusion on a dose response relationship based on efficacy results of the dose ranges. An efficacy comparison of dose ranges may not be meaningful when individual doses were optimized. At any rate, the efficacy estimates of the three dose ranges (3-6 mg, 6-12 mg and 3-12 mg) did not show a consistent dose response relationship.

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

/s/

EIJI ISHIDA
07/22/2013

PEILING YANG
07/22/2013
I concur with the review.

HSIEN MING J HUNG
07/22/2013



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

CARCINOGENICITY STUDIES

NDA: 204370 (IND 71958)

Drug Name: Cariprazine (RGH-188 HCl)

Indications: Treatment of 1) schizophrenia 2) manic and mixed episodes associated with bipolar disorder

Applicants: Sponsor: Forest Research Institute
Harborside Financial Center
Jersey City, New Jersey

CRO Rats: (b) (4)

CRO Mice: (b) (4)

Date: NDA Submitted: 19 November 2012

Review Priority: Standard

Biometrics Division: Division 6

Statistical Reviewer: Steve Thomson

Concurring Reviewers: Karl Lin, Ph.D.

Medical Division: Psychiatry Products

Toxicologist Team: Elzbieta Chalecka-Franaszek, Ph.D.
Aisar Atrakchi, Ph.D.

Project Manager: Kimberly Updegraff, MS

Keywords: Bayesian analysis, Carcinogenicity, Cox regresson, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY	3
1.1. CONCLUSIONS AND RECOMMENDATIONS	3
1.2. BRIEF OVERVIEW OF THE STUDIES	8
1.3. STATISTICAL ISSUES AND FINDINGS	8
1.3.1. <i>Statistical Issues</i>	8
1.3.2. <i>Statistical Findings</i>	15
2. INTRODUCTION	15
2.1. OVERVIEW.....	15
2.2. DATA SOURCES	15
3. STATISTICAL EVALUATION	15
3.1. EVALUATION OF EFFICACY.....	15
3.2. EVALUATION OF SAFETY	15
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	27
5. SUMMARY AND CONCLUSIONS	27
5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	27
5.2. CONCLUSIONS AND RECOMMENDATION	27
APPENDICES.....	28
APPENDIX 1. FDA SURVIVAL ANALYSIS	28
APPENDIX 2. FDA NONPARAMETRIC BAYESIAN SURVIVAL ANALYSIS.....	32
APPENDIX 3. FDA POLY-K TUMORIGENICITY ANALYSIS	41
APPENDIX 4. REFERENCES.....	66

1. EXECUTIVE SUMMARY

Reports and data from two studies, a two year study in standard rats and a six month study in transgenic mice, were provided. The rat study was conducted by (b) (4). The transgenic mice study, in Tg.rasH2 mice, was conducted at the (b) (4). The Sponsor states that: “The purpose of this study was to assess the carcinogenicity of the RGH-188 HCl, [i.e. Cariprazine,] an atypical antipsychotic, when administered orally to rats for most of their life span (2 years).” (page 15 of rat report). Note these studies were originally submitted under IND 71958. The final report for the rat study was submitted in November 18, 2011, and for the transgenic mouse study June 19, 2012.

1.1. Conclusions and Recommendations

In each study the drug vehicle is water, described as distilled in the rat study, sterile in the mouse study. For each study, in each gender, there are three actual treatment groups. Animals were dosed once daily by oral gavage. Gross aspects of the study designs for the main study animals are summarized in Tables 1 and 2 below:

Table 1. Design of Rat Study (dose volume 10 mL/kg)

Treatment Group	# Main study animals (# TK ^a animals)/gender	Male Dose (mg/kg/day)	Male Dosing Concentration (mg/mL)	Female Dose (mg/kg/day)	Female Dosing Concentration (mg/mL)
1. Vehicle ^b	60 (6)	0	0	0	0
2. Vehicle ^b	60 (6)	0	0	0	0
3. Low	60 (27)	0.25	0.027	1	0.109
4. Medium	60 (27)	0.75	0.082	2.5	0.273
5. High	60 (27)	2.5	0.273	7.5	0.818

^a Toxicokinetic phase animals began dosing during Week 1 of the carcinogenicity phase and terminated during Week 52

^b Distilled water alone.

In the rat study, the Sponsor notes that: “A 2-week dose adaptation period preceded the formal 2 year study. Animals were administered 1/3 of the final freebase dose for one week and 2/3 of the final freebase dose for 1 week prior to full dosing for up to 2 years.” (page 9 of rat report) In each treatment group in each gender a further 10 animals were specified as a clinical pathology group.

Table 2, below, provides a similar outline of the study design in the six month Tg.rasH2 transgenic mouse study. More detailed descriptions of the studies are provided in Sections 3.2.1 and 3.2.2 below. In each study in each gender, it is convenient to label the three actual Cariprazine dose groups as the “Low”, “Medium,” and “High” dose groups. As suggested in the table below, in the Sponsor’s report for the Tg.rasH2 study dose groups these groups are identified as groups 3 through 8, depending upon animal gender, where the odd numbered groups correspond to male dose groups and the even numbers to female dose groups. With the sole

exception of Table 2 below, in this report these group numbers are recoded to 3-5 in both genders.

Table 2. Design of Tg.rasH2 Mouse Study (dose volume 5 mL/kg)

Treatment Group Numbers and Labels	# Main study animals (# TK ^a animals)/gender	Male Dose (mg/kg/day)	Target Male Dosing Concentration (mg/mL)	Female Dose (mg/kg/day)	Target Female Dosing Concentration (mg/mL)
1. Vehicle ^b	25 (10)	0	0	0	0
2. Positive Control ^c	15 -	0	100 ^c	0	100
3,4. Low	25 (68)	1	0.05	5	0.25
5,6. Medium	25 (68)	5	0.25	15	0.75
7,8. High	25 (68)	15	0.75	50	2.5

^a Toxicokinetic phase animals began dosing during Week 1 of the carcinogenicity phase and terminated during Week 52

^b Sterile water.

^c Urethane in 0.9% Sterile Saline

In Appendix 1, Figures A.1.1 and A.1.2, for rats, display Kaplan-Meier estimated survival curves for each study group for each gender. Simple summary life tables in mortality are presented in the study specific sections of this report, i.e., Tables 16 and 17, on page 20. Results of statistical tests on survival in rats are summarized in Table 3 below:

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over groups 1-5	0.0004	0.0003	0.1194	0.1684
Homogeneity over groups 1+2, 3-5	0.0002	0.0002	0.0654	0.0979
No Trend over dose groups 1+2, 3-5	< 0.0001	< 0.0001	0.1704	0.2480
No difference between groups 1+2 vs 5	< 0.0001	< 0.0001	0.1175	0.2080
No difference between groups 1 & 2	0.3645	0.1993	0.7820	0.7317

From the Kaplan-Meier plot in Figure A.1.1 in Appendix 1, it seems that in male rats the high dose group tends to have the highest survival (i.e., lowest mortality) and the vehicle control group 1 generally with the the lowest survival. Through most of the study the remaining study groups are more or less intertwined until near the end of the study, when the mortality in the vehicle control group 2 increases to nearly match that in vehicle control group 1. As discussed in Section 1.3.1,1 below, this reviewer would argue that these apparent differences between the two essentially equivalent controls are likely due to the sort of random fluctuations that occur in any study and should be ignored. However the differences in survival (in a negative direction) between the pooled vehicle group and the actual Cariprazine dose groups are sufficient to result in highly statistically significant tests of overall differences in survival among the three study

groups and the pooled vehicle control (both logrank and Wilcoxon $p = 0.0002$). Similarly, tests of trend and pairwise differences between the high dose and vehicle were highly statistically significant (all four logrank and Wilcoxon $p < 0.0001$), though in a negative direction.

The Kaplan-Meier plot for female rats in Figure A.1.2 in Appendix 1 is even simpler to interpret, with vehicle control groups 1 & 2 having the lowest survival, and the other study groups largely intertwined. These differences are sufficient to result in tests of lack of homogeneity of the three treatment groups and pooled control that are close to statistical significance at the usual 0.05 level (logrank $p = 0.0654$, Wilcoxon $p = 0.0979$), although again in a negative direction. No other test of trend or treatment group differences were statistically significant (all other eight $p \geq 0.1175$).

An alternative Bayesian analysis of survival in the rat study using an accelerated failure time (AFT) model is presented in Appendix 2. In male rats its results are generally similar to those in the analysis reported above, but with the stipulation that one has estimates of actual probabilities that parameters have an effect of interest. In female rats the estimation method has problems that bring these results into question.

Although of arguably limited use, Kaplan-Meier survival plots for mice are presented in figures A.1.3 and A.1.4 in Appendix 1. The asymptotic tests used in the rat study are most appropriate when there are a reasonably large number of events (i.e., deaths). However, except for the positive control group there were few deaths in the transgenic Tg.rasH2 mouse study. That makes interpreting p-values somewhat problematic. An alternative view is presented in the table below, displaying the actual observed times to death of the animals in each of the main study groups. The day of terminal sacrifice and the number of animals sacrificed are listed to the right under each gender.

Table 4. Survival times of Tg.rasH2 mice.

Dose group	Male Mice			Female Mice		
	Day of death	Term. Sac.		Day of death	Term. Sac.	
		Day	#		Day	#
1. Vehicle	97,142	197	23	78,165	197	23
2.Pos Ctrl	31, 78,107,107,116,118	121	9	106,118,119,119	123	11
3.Low	151,198	198	23	119,197	198	23
4.Medium	55	198	24	94,121,142	198	22
5.High	-	197	25	170	197	25

Except for the positive control, there is little difference in death rates between study groups. In particular, it is apparent that in both genders there was little difference in survival in groups 1, 3, 4, and 5, and what little difference was manifest suggested a positive dose effect on survival. Statistical significance levels of exact logrank tests based on permutation distributions are provided in Appendix 1, but it was felt that these tests were not needed to conclude that, except for the positive control, there was little dose effect on survival in either gender in mice.

Of course in a carcinogenicity study, primary interest is on the occurrence of cancers. The statistical analysis of tumors compares tumor incidence over dose groups. The poly-k test, as used here with $k=3$, modifies the original Cochran-Armitage test to adjust for differences in mortality (please see section 1.3.1.4 for details). Complete tumor incidence tables for each organ tumor combination listed by the Sponsor in the submitted data sets and those combined by this reviewer are provided in Tables A.3.3 through A.3.7 in Appendix 3.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman (HLR) rules discussed in Section 1.3.1.5 are often applied. For tests of positive trend in dose in the two year study, common tumors should be tested at a 0.005 significance level and rare tumors at a 0.025 level, while for pairwise tests between the high dose and control, it is recommended that common tumors be tested at a 0.01 level while rare tumors be tested at a 0.05 level. In the six month transgenic study tests of trend and the tests of differences between the high dose and control are all tested at a 0.05 level. This approach essentially weights the studies separately. Using these adjustments for other tests, like testing the comparisons between the low and medium dose groups versus the vehicle or testing against the water group can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than that approximate overall 10% rate.

Table 5, below, shows those tumors in rats that had at least one mortality adjusted statistical test significant at a nominal 0.10 level. The only tumors in mice that met this criterion were the tests involving the positive control. Note that when one adjusts for multiplicity these nominally significant comparisons may not be statistically significant. Table A.3.3 in Appendix 3 displays the results of the revised Pathology Working Group (PWG) tumor analysis while Tables A.3.4-A.3.7 in this appendix display all incidences and statistical test results in the original data for both genders in rats and mice.

For each gender by organ combination, the number of animals supposedly microscopically analyzed is presented first. The entry for each tumor is preceded by the adjusted number of animals at risk for that endpoint. It seems clear that an animal that dies early without having displaying that endpoint reduces the size of the risk set for getting that particular endpoint. The poly-k test down weights such animals, and as discussed in Section 1.3.1.4, below, the sum of these poly-k weights seems to be a better estimate of the number of animals at risk of getting that tumor. This sum is given in the row labeled "Adjusted # at risk". Tumor incidence is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high and medium dose groups versus pooled vehicle. The next row continues with the p-values of the pairwise test between the low and pooled vehicle dose group. For these analyses, incidence in the pooled vehicle group is used to assess background tumor incidence, and thus whether a tumor is considered to be rare (background incidence $<1\%$) or common. Note that for this analysis a tumor is only classified as rare if the pooled vehicle group shows at most a single example of that particular tumor. Thus the only tumor in Table 5 below that would be classified as a rare tumor was schwannoma of the vagina in female rats. Finally, note that there are two sets of analyses for adrenals, one from the tumor specification of the

original toxicologist and another from the Pathology Working Group (PWG), a group of experts with expertise in rodent toxicity and carcinogenicity (please see Section 3.2.1 for details).

Table 5. Potentially Statistically Significant Neoplasms in Rats

Organ/ Tumor	Overall Results					ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
	Veh1	Veh2	Low	Med	High			
Male Rats								
PANCREAS								
# Evaluated	60	59	60	60	60			
Adj. # at Risk	35.4	38.7	39.6	43.4	50.1			
ISLET CELL ADENOMA	2	1	4	3	6	.0940	.0946	.3871
						.1850		
Adj. # at Risk	35.4	39.1	39.6	43.4	50.2			
Islet Cell Adenoma/Carcinoma	3	3	5	6	9	.0732	.0854	.2424
						.3122		
Female Rats								
ADRENAL GLANDS (PWG Analysis)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	46.6	45.9	49.4	50.8	48.9			
Medulla Pheochromocytoma [B&M]	1	3	1	1	8	.0020	.0178	.8902
						.8862		
ADRENAL GLANDS (Original Toxicologist)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.9	37.9	41.1	45.2	41.8			
MEDULLA BENIGN PHEOCHROMO - CYTOMA	1	1	0	1	4	.0226	.1119	.7559
						1		
Adj. # at Risk	38.9	37.9	41.1	45.2	42.8			
Medulla Pheochromocytoma [B&M]	1	1	0	1	5	.0072	.0541	.7559
						1		
PANCREAS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	38.4	42.1	45.6	41.7			
Islet Cell Adenoma/Carcinoma	0	3	5	2	2	.5700	.5700	.6109
						.1017		
SKIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.5	38.7	41.7	46.9	42.4			
FIBROMA	3	1	1	3	5	.0573	.1629	.5177
						.8844		
VAGINA								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	42.8			
SCHWANNOMA	0	0	0	0	2	.0416	.1247	.

Using the HLR rules to adjust for multiplicity, note that the test of trend in pooled benign and malignant pheochromocytoma of the adrenals was statistically significant in the PWG analysis ($p = 0.002 < 0.005$) and close to significance using the tumor attributed by the original

toxicologist ($p = 0.0072 \approx 0.005$). No other comparisons achieved the multiplicity adjusted levels of statistical significance for a single study.

Except for the tests involving the positive control, no tumors in mice achieved the even a 0.10 significance level. Those tests are listed in table 22 in Section 3.2.2. The results of all tests in Tg.rasH2 mice are also presented in Appendix 3.

1.2. Brief Overview of the Studies

This submission had a rat study RGH-TX-34:

RGH-188 HCl: 2-Year Oral (Gavage) Carcinogenicity Study in Rats,

conducted at [REDACTED] ^{(b) (4)} and a transgenic mouse study RGH-TX-33 :

Cariprazine: 28-Week Repeated Dose Oral Carcinogenicity Study in Tg.rasH2 Mice,

conducted at [REDACTED] ^{(b) (4)}. Fairly detailed descriptions of these studies are available in Sections 3.2.1 and 3.2.2, below.

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 details on the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Control Groups:

In the rat study, the Sponsor provides two supposedly identical vehicle control groups. All tables and plots in this report distinguish between the two vehicle control groups, vehicle groups 1 and 2. Prior to tests the Sponsor tests for differences in the control groups and provides different tests on the basis of these tests. The first issue with such a 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, unless there are systemic differences in the controls, 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 such post study observed differences between the control groups (since a simple rerandomization would likely 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 conducted by the Sponsor, and although tests differentiating between control groups are provided, this reviewer would recommend they be ignored. In the formal FDA analysis, both tests of differences in survival and tests of differences tumorigenicity use a single pooled control group and ignore the possible differences in controls.

Note that in a two year study the determination of whether or not a tumor is classified as rare or common has a considerable impact upon results. Given the lack of good historical data in a genetically stable line of animals it seems that in both studies the vehicle group should be used to make this determination.

1.3.1.2. Survival Analysis:

The survival analyses in rats presented here are based on both the log rank test and the Wilcoxon test comparing survival curves. The log rank tests tend to put 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 logrank test is most powerful when the survival curves track each other, and thus the hazards, i.e., the conditional probability of the event in the next infinitesimal interval, would be roughly proportional. Note the logrank test seems to be the test usually recommended by statisticians. Both tests are use in the FDA analysis of mortality in rats. The Sponsor provides results from a Fisher exact test in the mouse study, while FDA statistical tests on mortality are based on exact version of the logrank test. Arguably, neither test is needed in the mouse study. Appendix 1 reviews the specific FDA animal survival analyses in more detail. The results of the Sponsor's analysis are summarized in Sections 3.2.1.1 and 3.2.2.1.

1.3.1.3. Multiplicity of Tests on Survival:

Using both the logrank and Wilcoxon tests, for each gender in rats there are 10 tests of survival differences. Assuming tests were performed at the usual 0.05 level, and the tests were stochastically independent, but there were actually absolutely no differences in survival across groups (so one would hope no tests would be statistically significant), the probability of at least one statistically significant result in each species by gender was about 0.40. These bounds assume the tests are independent, which they clearly are not, but these values can give some idea of the possible price paid for the multiplicity of hypothesis tests in the statistical frequentist paradigm.

1.3.1.4. Tests on Neoplasms:

The data sets requested for the analysis of rodent carcinogenicity studies are supposed to include a record for each animal organ combination that was not evaluated. If a number of the animals are not examined, but the proportions of animals showing the tumor under study in each treatment group is roughly the same as in the subset of animals actually reported the calculated p-values will generally be too large, i.e., results will be less statistically significant than they should be, possibly much less. If we can assume the process that determines whether or not a

tumor is analyzed in each specific tumor is random, it is perhaps appropriate to consider such endpoints to be both analyzed AND have the tumor.

Ignoring these possible problems, the Sponsor's analyses of tumorigenicity are Peto tests (Peto *et al*, 1980) for trend for those tumors with an incidence of at least two in the three treated groups. These require accurate determination of whether a tumor is fatal or incidental.

The FDA analysis is based on a modification of the Cochran-Armitage test of trend in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). Inspecting a large number of studies, Bailer and Portier noted that survival time seemed to fit a Weibull distribution, generally with a shape parameter of between 1 and 5, with 3 a typical value. With t_{\max} denoting the maximal time to terminal sacrifice and t_{obs} the time to death of the animal, they proposed weighting the animal by $(t_{\text{obs}}/t_{\max})^k$, so that an animal that survives for say 52 weeks in 104 week study without the tumor being analyzed is counted as $(1/2)^k$ of an animal. For $k = 3$, that means that particular animal would count as 1/8 of an animal in the analysis of that tumor. Further, the $k = 3$ specification seems to represent tumor incidence where some animals are perhaps more sensitive and respond earlier to the insult than the remaining animals. Under this structure time to incidence would tend to follow a cubic expression. Thus an animal with the specific tumor being studied or who survives to terminal sacrifice without the tumor will be given a weight of 1 when counting the number of animals at risk. However, animals that die early without the tumor are down weighted when counting the number of animals in the risk set for that specific tumor. With differential mortality, this can mean a substantial reduction in the size of that risk set. Note this seems to be an appropriate adjustment for dose groups that are terminated early as in the rats study. The report of the Society of Toxicological Pathology "town hall" meeting in June 2001 recommended the use of this poly-k modification of the so-called Cochran-Armitage tests of trend over the corresponding Peto tests used by the Sponsor.

The computed significance levels are based on small sample exact permutation tests of tumor incidence. In the tumor incidence tables the effective size of the risk set for each tumor is listed in the row labeled "Adjusted # at risk", and seems to be a more appropriate denominator when comparing incidence rates than the simple unadjusted number evaluated.

1.3.1.5. Multiplicity of Tests on Neoplasms:

Frequentist hypothesis testing involves accepting or rejecting hypotheses about the parameters of interest on the basis of the values of some statistic. If one does not provide some sort of multiplicity adjustment to the significance level, the chances of rejecting one or more true null hypothesis increases as the number of such tests increases. To avoid this, it is common to adjust for multiplicity in hypothesis testing resulting in an adjustment in experiment-wise Type I error (i.e., the probability of rejecting a true null hypothesis and thus concluding there is an effect when in fact there is none). Based on his extensive experience with such carcinogenicity analyses in standard laboratory rodents, for pairwise tests between the highest 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,

simulations by Lin and Rahman (1998) indicated that tests of trend over all dose groups should be tested at about a 0.025 level for rare tumors and 0.005 for common tumors.

These rules apply to standard two year studies in two species, but transgenic mouse studies are of shorter duration with less time for tumor development. Based on further experience and simulations these recommendation have been revised so that if one were only interested in tests of positive trend in dose, in the two year study common tumors should be tested at a 0.005 significance level and rare tumors at a 0.025 level, while both common and rare tumors would be tested at a 0.05 level in the shorter transgenic mouse study. For the shorter study no distinction is made between the trend and the high dose and control, or between common and rare tumors. It is recommended that all should be tested at 0.05 level.

“For tests for positive trend and control-high pairwise comparison jointly, it is recommended that common and rare tumors are tested at 0.005 and 0.025 significance levels, respectively in trend test, and at 0.05 and 0.10 significance levels, respectively, in control-high pairwise comparison in the two-year study; and at 0.05 and 0.05 significance levels, respectively, in both trend test and control-high pairwise comparison in the alternative study.

“The use of the above decision rules in a submission under the ICH guideline will result in an overall false positive rate about 5 % in the two-year study and another overall false positive rate also around 5% for the short- or medium-term studies in the tests for trend and control-high groups pairwise comparisons, separately or jointly. This will result in an overall false positive rate around 10% for the entire submission.” (page 32 of 2013 draft guidance)

Note that significance levels of the pairwise tests between the water vehicle, and Low and Medium dose groups, plus, in mice, a comparison of the vehicle and positive control group are also provided. Even following the HLR rules, adding these comparisons can be expected to increase the overall type I error rate to some level above the usual rough 10% level, possibly considerably larger. Again, because of the possibility of genetic drift and for convenience the vehicle group is used to determine if the tumor is classified as rare or common.

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 (i.e. 50%), 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. As a percentage of the High dose group animals that survived to week 91, this criterion does seem to be met in rats (High dose: Males 73.3%,

Females 70%, both > 50%). In fact this criterion is met in all actual dose groups save the low dose in males (but 48.3% ≈ 50%). ..

The mean weight values used to derive differences and ratios in the following tables were taken directly from the Sponsor's reports (in the rat study, Table 5, Mean Body Weights Values, pages 100-119, and Table 6, Mean Body Weight Changes from Baseline, pages 121-140, in the mice study, Table 11-5, Summary of Body Weights, pages 67-72, and Table 11-6 Summary of Body Weight Gains, pages 73-80) S in rats, and pages 67-80 in mice). The change from baseline in the table below is the simple difference between the means at the specified dates, and thus animals that die early are only counted at the study initiation, not at the end of the study.

Table 6. Mean Weights and Changes (in g) in Male Rats

Dose Group	Dose mg/kg/day	Week		Change from Baseline	% change relative to vehicle
		Pre-Test 1	99		
1. & 2. Vehicle	0	115.5	721.7	606.2	
3. Low	0.25	115.7	815.0	699.3	115.7%
4. Medium	0.75	115.4	725.1	609.7	100.6%
5. High	2.5	113.8	664.7	550.9	90.9%

Table 7. Mean Weights and Changes (in g) in Female Rats

Dose Group	Dose mg/kg/day	Week		Change from Baseline	% change relative to vehicle
		Pre-Test 1	105		
1. & 2. Vehicle	0	100.5	567.4	466.9	
3. Low	1	100.3	587.9	487.6	104.4%
4. Medium	2.5	100.4	554.6	454.2	97.3%
5. High	7.5	100.4	406.4	306.2	65.6%

Table 8. Mean Weights and Changes (in g) in Male Tg.rasH2 Mice

Dose Group	Dose mg/kg/day	Day			Change from Baseline/Day		% change relative to vehicle/Day	
		1	113	197	113	197	113	197
1. Vehicle	0	22.69	27.60	29.30	4.91	6.61		
2. Pos. Control	100	22.81	28.97	NA	6.16	NA	125.4%	NA
3. Low	1	22.74	27.17	28.29	4.43	5.55	90.2%	84.0%
4. Medium	5	22.92	25.36	26.39	2.44	3.47	49.7%	52.5%
5. High	15	22.42	24.87	25.97	2.45	3.55	49.9%	53.7%

Table 9. Mean Weights and Changes (in g) Female Tg.rasH2 Mice

Dose Group	Dose mg/kg/day	Day			Change from Baseline/Day		% change relative to vehicle/Day	
		1	113	197	113	197	113	197
1. Vehicle	0	18.10	21.14	22.47	3.04	3.37		
2. Pos. Control	100	18.11	22.32	NA	4.22	NA	133.8%	NA
3. Low	5	18.28	22.99	23.40	4.71	5.12	154.9%	151.9%
4. Medium	15	18.08	21.78	22.66	3.70	4.42	121.7%	131.2%
5. High	50	17.81	20.90	21.96	3.09	4.15	101.6%	123.1%

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 Tables 6 and 7 above, the weight decrement criterion is clearly exceeded in the high dose in female rats but not in the nominal high dose in male rats. It is not clear if the same criterion is applicable to transgenic mice, but, interestingly, in Tg.rasH2 mice the situation is reversed. The high dose is associated with a considerable weight reduction in Tg.rasH2 male mice and an increase in females. . Again, although this requires the expertise of the toxicologist, this may be evidence that the MTD was exceeded in female rats and possibly in male Tg.rasH2 mice.

More generally, in the rat study, the Sponsor notes that: “RGH-188 HCl administration resulted in dose-related, statistically significant ($p \leq 0.01$) decreases in mean body weight gain and mean body weights in males at all doses (0.25, 0.75 and 2.5 mg/kg/day) and in females at the highest dose (7.5 mg/kg/day). Differences from the pooled mean control body weight were generally slight in the males at 0.25 and 0.75 mg/kg/day (up to 6 or 8%) respectively. At the high dose, differences from control were up to 16% in males and 21% in females. Body weight gains and mean body weights for females at 2.5 mg/kg/day were higher than concurrent control values throughout the study and values for females at 1 mg/kg/day were higher than control values for the first year of study and similar or lower than control subsequently.” (page 48 of rat report)

The Sponsor provides a detailed comparison of body weights in the mouse study. To summarize, there was a general group mean body weights decrease over increasing dose, although the low dose group was generally comparable to the vehicle control. However the effects in females were not as striking as with males.

In a related issue the Sponsor notes that in the rat study: “Food consumption for all RGH-188 HCl-treated groups was lower than concurrent control values throughout the study; differences were generally dose-related and statistically significant. These differences are consistent with body weight patterns for all treated groups of males and for the high dose females.” (page 48-49 of rat report)

In the Tg.rasH2 mouse study the Sponsor reports: “Statistical analysis of weekly food consumption data revealed sporadic statistically significant differences (both increases and decreases) in both sexes when the test article treatment groups and positive control were compared to the vehicle control. These differences were few in number, did not occur in a dose-related manner and were not considered related to treatment with the test or positive control articles.” (page 30 of mouse report)

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. This suggests that a useful way to assess whether or not the MTD was achieved is to measure early mortality not associated with any identified tumor. If this is high in the higher dose groups it suggests that animals tend to die before having time to develop tumors. Tables 10 and 11, below, display 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 10. Natural Death with No Identified Tumor in Rats (Male/Female)

		1. Vehicle	2. Vehicle	3. Low	4. Medium	5. High
Males	Event	16	7	12	10	5
	No event	44	53	48	50	55
Females	Event	2	1	1	1	2
	No event	58	59	59	59	58

It is apparent that there is no evidence of heterogeneity in dying without tumor in female rats. This is confirmed in using a Fisher exact test of a lack of homogeneity ($p = 0.6063$). Although differences in males seem to be more apparent, but these differences are not statistically significant (chi square $p = 0.3574$, Fisher exact $p = 0.3401$). So neither gender seems to show dose related treatment differences in excess mortality unrelated to tumorigenicity.

Table 11. Natural Death with No Identified Tumor in Mice (Male/Female)

		1. Vehicle	2. Positive Control	3. Low	4. Medium	5. High
Males	Event	1	0	1	0	0
	No event	24	15	24	25	25
Females	Event	0	0	2	1	1
	No event	25	15	23	24	24

In mice note that there is no particular evidence of any dose related differences in excess mortality in either gender, let alone an increasing dose related effect. Thus, there is no evidence of excess mortality unrelated to tumorigenicity in either gender in either rodent species. Like the other observations above, this require the expertise of the toxicologist, but these tests may provide evidence that the MTD was not exceeded in either gender in either rats or mice.

Combining these perhaps somewhat inconsistent observations into a valid assessment of whether or not the MTD was exceeded requires the expertise of the toxicologist.

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 and a shorter term transgenic mouse Tg.rasH2 study to assess the carcinogenic potential of the test article.

2.2. Data Sources

The original data was provided in two SAS transport files: “tumour.xpt” for rats and “tumor.xpt” for mice, each containing a SAS data set labeled “tumour.sas7bdat” and “tumor.sas7bdat,” respectively. In addition a small data set containing revised tumor assessments of neoplasms in the adrenals of rats by the Pathology Working Group (PWG) was also provided. Several inconsistencies in the TUMERCOD and TUMORNAM variables in this data set were resolved to be consistent with the tumor name. Note these data sets contained sufficient survival and tumorigenicity data to conduct the primary analyses in this report. Certain tumors and organs were combined for analysis following the recommendation of the toxicologist.

Note that the tumor detection time is used as proxy for the time of tumor development in the tumor analysis. For some animals the time of detection may be earlier than the time of death.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1. RGH-188 HCl: 2-Year Oral (Gavage) Carcinogenicity Study in Rats,

STUDY DURATION: Males 23 months, Females 24 months

STARTING DOSING DATE: Dose Adaption: 24 August 2006

Full Dosing Started: 7 September 2006

FINAL DOSING DATES: Males: 27-31 July 2008

Females: 7-11 September 2008

TERMINAL SACRIFICE DATES: Males; 27 July 2008 – 1 August 2008

Females 8-12 September 2008

RAT STRAIN: Wistar Albino Rats (Outbred) CrI: (WI)BR

ROUTE: Daily Oral Gavage

The Sponsor states that for this study: “Dose levels were based on the results of a 13-week oral toxicity study in rats ...and followed input from the FDA Carcinogenicity Assessment

NDA 204370 Cariprazine Forest Research Institute
 Committee (CAC).” (page 19 of rat report). The study design is summarized in Table 12. below
 (a repeat of Table 1 with group numbers recoded):

Table 12. Design of Rat Study (dose volume 10 mL/kg)

Treatment Group	# Main study animals (# TK ^a animals)/gender	Male Dose (mg/kg/Day)	Male Dosing Concentration (mL/kg)	Female Dose (mg/kg/day)	Female Dosing Concentration (mL/kg)
1. Vehicle ^b	60 (6)	0	0	0	0
2. Vehicle ^b	60 (6)	0	0	0	0
3. Low	60 (27)	0.25	0.027	1	0.109
4. Medium	60 (27)	0.75	0.082	2.5	0.273
5. High	60 (27)	2.5	0.273	7.5	0.818

^a Toxicokinetic phase animals began dosing during Week 1 of the carcinogenicity phase and terminated during Week 52

^b Distilled water alone.

The Sponsor notes that “A 2-week dose adaptation period preceded the formal 2 year study. Animals were administered 1/3 of the final freebase dose for one week and 2/3 of the final freebase dose for 1 week prior to full dosing for up to 2 years.” (page 9 of rat report) The Sponsor notes that “Due to declining survival in the two control groups of males, the males were terminated after approximately 23 months at the full dose. The females were terminated as scheduled after 24 months of treatment. At the end of the treatment periods, all surviving animals were euthanized and necropsied.” (page 7 of rat report)

Randomization to treatment was stratified. As the Sponsor explains “Animals considered suitable for study were distributed into 2 groups (controls) of 76 animals per sex or 3 groups (treated) of 97 animals per sex by a computerized random sort program so that body weight means for each group were comparable. Individual weights of animals placed on test were within ±20% of the mean weight for each sex. Information as to the disposition of all animals not utilized in the study is maintained in the study file.” (page 23 of rat report)

“Animals were doubly housed during the first week of the acclimation period and individually housed thereafter in elevated, stainless steel, wire mesh cages or other appropriate cages.” (page 24 of rat report) The Sponsor reports that animals were dosed daily, and were after initiation of dosing were housed individually with food and water was available *ad libitum* throughout the study.

Note there are two sets of analyses of adrenals. As explained by the toxicologist reviewer: “Apparently, after submission of the final report of the rat carcinogenicity study (conducted by the (b)(4)) to the Agency for review for the first time (in November 2011, under IND 71958), the incidence of proliferative changes involving the adrenal medulla was reevaluated by the Pathology Working Group (PWG), a group of experts with expertise in rodent toxicity and carcinogenicity led by Dr. (b)(4) .,. Their report, dated August 29, 2012, is attached to the rat

carcinogenicity study report submitted on November 19, 2012, as a part of the NDA submission.” (personal communication) Note that the incidence of the adrenal medulla tumors reported by the PWG differs from that of the original toxicologist. Both sets are analyzed in this 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 summarized results as follows: “Mortality in test article-treated groups was lower than in control groups for both males and females. The difference was dose-related and statistically significant ($p < 0.001$) for males at 2.5 mg/kg/day. A trend test demonstrated significance for the males at 0.75 mg/kg/day but there was no significance in pair-wise comparisons for this group or for males at 0.25 mg/kg/day. There were no statistically significant trends or differences from control for females at any dose. A similar trend in mortality was evident in the Clinical Pathology satellite subgroup that was treated throughout the study but not included in mortality analysis. This decreased mortality is consistent with the decreased body weight gains and consequent low body weights in all treated groups of males and the females treated at 7.5 mg/kg/day. Low body weight gains consequent to dietary restriction have been shown to enhance survival in rats. . . .

“Because of declining survival in the two control groups of males, all males were terminated at Month 23. It should be noted that, because of the 2-week dose adaptation period, males received a total of 100 to 101 weeks of test article administration. Females were terminated as scheduled, at the end of 2 years (24 months).” (pages 46-47 of rat report) The Sponsor summarized results in the following (slightly edited table).

Table 13. Sponsor Table 3.3-1: Mortality and Cause of Death Summary

Dose (mg/kg/day):	Males					Females				
	0 ^a	0 ^a	0.25	0.75	2.5	0 ^a	0 ^a	1	2.5	7.5
Total Number	60	60	60	60	60	60	60	60	60	60
No. Preterm. Deaths	45	43	34	34	25	33	34	24	23	25
No. Survivors	15	17	26	26	35	26	26	36	37	35
Percent Mortality	75%	72%	57%	57%	42%	57%	57%	42%	38%	42%
Cause of Death:										
Pituitary Neoplasm	4	6	3	1	2	13	16	9	4	4
Mammary Neoplasm	0	0	0	0	1	7	5	8	8	5
Chronic Nephropathy	14	17	13	6	9	3	2	1	0	0
Skin Neoplasms	4	3	4	2	2	0	1	1	3	4
Undetermined	10	3	6	10	3	1	4	1	1	3
Other	13	14	8	15	8	9	6	4	7	9

^a Vehicle treatment

. (pages 46-47 of rat report)

Tumorigenicity analysis:

The Sponsor's report describes Peto analyses of tumorigenicity where the results on fatal and incidental tumors were pooled, and mortality independent tumors were pooled with a life table analysis of incidental tumors. "For non-incidental tumours, the strata are defined as those weeks during which there were deaths. For incidental tumours, the following fixed time intervals were used to adjust for differential mortality between the treatment groups: For males 1-52, 53-78, 79-92, 93-98 weeks and terminal sacrifice and for females 1-52, 53-78, 79-92, 93-104 weeks and terminal." (page 8 of statistical report on tumorigenicity, page 436 of overall rat report)

The Sponsor summarizes carcinogenicity (i.e., neoplastic) results as follows: "There were no statistically significant increases in the incidence of tumors in RGH-188 HCl-treated males.

"In females there was a statistically significant increase in benign tumors of the adrenal medulla (pheochromocytoma) at 7.5 mg/kg/day when compared to controls. Trend analysis also indicated slight increases in the incidence of fibroma in the skin and schwannoma of the vagina. Only pheochromocytoma in adrenal medulla at the 7.5 mg/kg/day dose level was considered to be related to RGH-188 HCl administration.

"For benign and malignant medullary pheochromocytoma of the adrenal combined, the trend test was statistically significant ($p=0.005$) across test article-treated groups when compared with the pooled controls for females. Pairwise comparisons of the test article-treated groups to the pooled control group was [sic] statistically significant ($p=0.048$) only for the 7.5 mg/kg/day females.

"The trend test for benign medullary pheochromocytoma was statistically significant across groups ($p=0.015$) but no test article-treated groups were statistically significant by pairwise comparison with the pooled control group. Pheochromocytomas have been described in rats following the administration of many different non-genotoxic xenobiotics including neuroleptic drugs (Greaves, 2007). However, it has been argued that the sensitivity of the rat adrenal medulla to the induction of pheochromocytoma by a wide variety of xenobiotics suggests that these lesions have little or no relevance to human safety when produced in rats following administration of high doses of non-genotoxic drugs (Greaves, 2007).

"For benign fibroma of the skin and malignant schwannoma of the vagina, the trend test was statistically significant across test article-treated groups when compared with the pooled controls ($p=0.047$ and $p=0.038$, respectively) but there were no statistically significant differences for any test article-treated group in pairwise comparisons to the pooled controls. In addition, there was no increase in the combined incidence of schwannomas from soft tissue at other sites in the body. This suggests that RGH-188 HCl was not associated with fibroma or malignant schwannoma formation in these organs." (pages 51-52 of rat report)

The following tables were copied from the Sponsors report:

Table 14. Sponsor Table 3.9.2-1: Incidence of Statistically Significant Tumors in Females

Test article (dose units)	0	0	1	2.5	7.5
# animals examined	60	60	60	60	60
Adrenals: B-pheochromocytoma	1	1	0	1	4 ^a
Adrenals: M-pheochromocytoma	0	0	0	0	1 ^a

^a Only the combined incidence, not the individual incidences, statistically significant (page 52 of report)

The Sponsor continues: “All other neoplasms occurred at similar incidences in control and test article-treated groups or they occurred sporadically with no dose relationship. The total number of benign and malignant neoplasms was similar between groups and the number of animals with one or more neoplasms was also similar between groups (Table 3.9.2-2).” (page 52 of rat report).

Table 15. Sponsor Table 3.9.2-2: Summary Incidence of Benign and Malignant Neoplasms

Dose Level (mg/kg):	Males					Females				
	0	0	0.25	0.75	2.5	0	0	1	2.5	7.5
Number of animals examined:	60	60	60	60	60	60	60	60	60	60
Total Benign Neoplasms	55	66	48	50	76	107	91	88	84	86
Animals with >1 Benign Neoplasms	32	43	32	36	46	49	49	47	50	42
Total Malignant Neoplasms	65	109	48	55	24	44	77	48	24	42
Animals with >1 Malignant Neoplasms	16	24	17	21	16	25	19	17	18	21

(page 53 of rat 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 16 for male rats, Table 17 for females) summarize the mortality results for the study 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. In these tables the terminal period only includes those animals were sacrificed. Rats that died of other causes during the terminal period are included in the preceding, overlapping time period. The Kaplan-Meier survival plots in Appendix 1 provide a more detailed picture of the profile of mortality losses.

Table 16. Summary of Male Rats Survival (dose label/dose¹/weeks dosing)

Period (Weeks)	Vehicle 1 0/1-105	Vehicle 2 0/ 1-105	Low 0.25/ 1-105	Medium 0.75/1- 105	High 1/1-105
1-52	11/60 ² 81.7% ³	4/60 93.3%	6/60 90.0%	3/60 95.0%	2/60 96.7%
53-78	13/49 60.0%	13/56 71.7%	13/54 68.3%	11/57 76.7%	5/58 88.3%
79-91	12/36 40.0%	15/43 46.7%	12/41 48.3%	13/46 55.0%	9/53 73.3%
92-104	9/24 25.0%	11/28 28.3%	3/29 43.3%	7/33 43.3%	9/44 58.3%
Terminal ⁴ 105	15	17	26	26	35

¹ dose in mg/kg/day² number of deaths / number at risk³ overall per cent survival to end of period.⁴ number of animals that survived to terminal sacrifice weeks**Table 17. Summary of Female Rats Survival (dose label/dose¹/weeks dosing)**

Period (Weeks)	Vehicle 1 0/1-105	Vehicle 2 0/ 1-105	Low 1/ 1-105	Medium 2.5/1- 105	High 7.5/1-105
1-52	2/60 ² 96.7% ³	1/60 98.3%	1/60 98.3%	2/60 96.7%	2/60 96.7%
53-78	7/58 85.0%	7/59 86.7%	6/59 88.3%	4/58 90.0%	9/58 81.7%
79-91	9/51 70.0%	14/52 63.3%	7/53 76.7%	8/54 76.7%	7/49 70.0%
92-102	15/42 45.0%	12/38 43.3%	10/46 60.0%	9/46 61.7%	7/42 58.3%
Terminal ⁴ 102,103	27	26	36	37	35

¹ dose in mg/kg/day² number of deaths / number at risk³ overall per cent survival to end of period.⁴ number of animals that survived to terminal sacrifice

Table 18 below provides the significance levels of the tests of homogeneity and trend over dose groups as proposed in Section 1.3.1.1 above (and is a repeat of Table 3 above and Table A.1.1 in Appendix 1). Recall that for statistical tests the nominally identical vehicle controls are pooled.

Table 18. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Rat Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over groups 1-5	0.0004	0.0003	0.1194	0.1684
Homogeneity over groups 1+2, 3-5	0.0002	0.0002	0.0654	0.0979
No Trend over dose groups 1+2, 3-5	< 0.0001	< 0.0001	0.1704	0.2480
No difference between groups 1+2 vs 5	< 0.0001	< 0.0001	0.1175	0.2080
No difference between groups 1 & 2	0.3645	0.1993	0.7820	0.7317

From the Kaplan-Meier plots in Figures A.1.1 and A.1.2 in Appendix 1, it seems that in both genders in rats the high dose group tends to generally have the highest survival, i.e. lowest mortality, while the two control groups more or less have the lowest survival, i.e. highest mortality, with the other dose groups in between these limits. However, these apparent differences are only statistically significant in male rats. The (not recommended) test of overall homogeneity in male rats was statistically significant (Logrank $p = 0.0004$, Wilcoxon $p = 0.0003$), while the test of homogeneity when the control groups are pooled was slightly more significant (Logrank $p = 0.0002$, Wilcoxon $p = 0.0002$), although both are in a negative direction, i.e. higher survival in the Cariprazine treatment groups. Similarly, in male rats the tests of trend and pairwise difference between the pooled controls and the high dose were all highly statistically significant (all four $p < 0.0001$). In female rats the test of homogeneity with control groups pooled was close to statistical significance (Logrank $p = 0.0654$, Wilcoxon $p = 0.0979$). No other test in female rats achieved even a 0.10 level of statistical significance, let alone the usual 0.05 level (all remaining $p \geq 0.1174$).

Tumorigenicity analysis:

As discussed in Section 1.3.1.5, the Haseman-Lin-Rahman rules for adjusting for multiplicity in a two year study with a short term 6 month study specify that for a very rough 0.10 (10%) overall false positive error rate it is common and rare tumors are tested at 0.005 and 0.025 significance levels, respectively in trend tests, and at 0.05 and 0.10 significance levels, respectively, in control-high pairwise comparisons

Table 19 below in rats show the tumors that had at least one mortality adjusted test whose nominal statistical significance was at least no more than 0.10. Note that when one adjusts for multiplicity these nominally significant comparisons most of these comparisons would not be statistically significant. Tables A.3.4-A.3.5 in Appendix 3 display all incidences and statistical test results for both genders in mice and rats. In this analysis we use the incidence in the sterile water vehicle control group to specify whether a tumor is treated as common or rare. Note that the period ‘.’ in the table denotes the p-values of tests of dose groups with none of the particular tumors the specified groups.

Table 19. Potentially Statistically Significant Neoplasms in Rats

Sex/ Organ/ Tumor	Incidence					Significance Level		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsveh
Male Rats								
PANCREAS								
# Evaluated	60	59	60	60	60			
Adj. # at Risk	35.4	38.7	39.6	43.4	50.1			
ISLET CELL ADENOMA	2	1	4	3	6	.0940	.0946	.3871
						.1850		
Adj. # at Risk	35.4	39.1	39.6	43.4	50.2			
Islet Cell Adenoma/Carcinoma	3	3	5	6	9	.0732	.0854	.2424
						.3122		
Female Rats								
ADRENAL GLANDS (PWG Analysis)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	46.6	45.9	49.4	50.8	48.9			
Medulla Pheochromocytoma [B&M]	1	3	1	1	8	.0020	.0178	.8902
						.8862		
ADRENAL GLANDS (Original Toxicologist)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.9	37.9	41.1	45.2	41.8			
MEDULLA BENIGN PHEOCHROMO- CYTOMA	1	1	0	1	4	.0226	.1119	.7559
						1		
Adj. # at Risk	38.9	37.9	41.1	45.2	42.8			
Medulla Pheochromocytoma [B&M]	1	1	0	1	5	.0072	.0541	.7559
						1		
PANCREAS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	38.4	42.1	45.6	41.7			
Islet Cell Adenoma/Carcinoma	0	3	5	2	2	.5700	.5700	.6109
						.1017		
SKIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.5	38.7	41.7	46.9	42.4			
FIBROMA	3	1	1	3	5	.0573	.1629	.5177
						.8844		
VAGINA								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	42.8			
SCHWANNOMA	0	0	0	0	2	.0416	.1247	.

Using the HLR rules to adjust for multiplicity, note that the test of trend in pooled benign pheochromocytoma of the adrenals was statistically significant in the PWG analysis ($p = 0.002 < 0.005$) and close to significance using the tumor attribution used by the original toxicologist ($p = 0.0072 \approx 0.005$). No other comparisons achieved the multiplicity adjusted levels of statistical significance for a single study.

Complete tumor incidence tables, including the adjusted number at risk, are provided in tables A.3.5 and A.3.6 of Appendix 2.

3.2.2. Cariprazine: 28-Week Repeated Dose Oral Carcinogenicity Study in Tg.rasH2 Mice

STUDY DURATION: 26 Weeks

STARTING DOSING DATE: Males: 20 October 2010

Females: 18 October 2010

FULL DOSING DATE: Males: 1 November 2010

Females: 3 November 2010

FINAL DOSING DATE: Males: 4-5 May 2012.

Females: 2-3 May 2012

MOUSE STRAINS: Main Study: “CByB6F1-Tg(HRAS)2Jic (+/- hemizygous c-Ha-ras mice)”,
(i.e. Tg.rasH2 Mice)

TK Study: “CByB6F1-Tg(HRAS)2Jic (-/- homozygous c-Ha-ras) (Tg.rasH2 nontransgenic littermate (referred to as CByB6F1)) is the same genetic background as the Tg.rasH2 mouse except for the omission of the Tg element”. (page 17 of mouse report)

ROUTE: Daily Oral Gavage

The Sponsor summarized the study as follows: “The purpose of this study was to assess the carcinogenic potential of cariprazine and to establish its toxicokinetic profile following repeated oral administration (gavage) for two weeks of dose escalation (mid-dose females and high-dose of both sexes) and 26 weeks of the final chronic dose in hemizygous Tg.rasH2 mice. The final full dose levels were 1 mg/kg/day (low-dose males), 5 mg/kg/day (mid-dose males and low-dose females), 15 mg/kg/day (high-dose males and mid-dose females) and 50 mg/kg/day (high-dose females). The escalation period of 2 weeks was to minimize signs of acute toxicity in the mid-dose females and high-dose animals of both sexes followed by administration of the full dose level from Weeks 3 through 28.

“ The Main Study consisted of three test article-treatment groups and one vehicle control group of twenty-five male and twenty-five female Tg.rasH2 transgenic mice each, plus one positive control group of fifteen male and fifteen female Tg.rasH2 transgenic mice. The TK Study consisted of three test article treatment groups of sixty-eight male and sixty-eight female wild type CByB6F1 mice each and one vehicle control group of ten animals per sex.” (page 8 of mouse report)

The study is summarized in Table 20 below (essentially a repeat of Table 2):

Table 20. Design of Tg.rasH2 Mouse Study (dose volume 5 mL/kg)

Sponsor Treatment Group Numbers and Labels (FDA renumbered 1-5)	# Main study animals (# TK ^a animals)/gender	Male Dose (mg/kg/day)	Target Male Dosing Concentration (mL/kg)	Female Dose (mg/kg/day)	Target Female Dosing Concentration (mL/kg)
1. Vehicle ^b	25 (10)	0	0	0	0
2. Positive Control ^c	15 -	0	100 ^c	0	100
3. Low	25 (68)	1	0.05	5	0.25
4. Medium	25 (68)	5	0.25	15	0.75
5. High	25 (68)	15	0.75	50	2.5

^a Toxicokinetic phase animals began dosing during Week 1 of the carcinogenicity phase and terminated during Week 52

^b Sterile water.

^c Urethane in 0.9% Sterile Saline

Animals were randomized to treatment balancing on mean weight. After randomization, animals were housed individually with food and water available *ad libitum*.

The Sponsor justified dosing as follows: “Doses used in this study were based on the results of two previous cariprazine studies conducted at [REDACTED]^{(b)(4)} with hybrid or transgenic mice. In study AB38JU.23GR.01 (RGH-TX-27), mice were dosed with 15, 20, and 30 mg/kg/day for up to six weeks. In study AB38JU.23GR.02 (RGH-TX-44), mice were treated with up to 60 mg/kg/day (males) or 140 mg/kg/day (females) for up to 7 weeks.” (page 19 of mouse 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 mice.

Survival Analysis:

“Analysis of the Main Study mortality data did not reveal any significant ($p < 0.05$) differences when the test article treated animals of either sex compared to the vehicle control. Two vehicle control (Group 1) Main Study males were sacrificed in a moribund condition (Days 97 and 142), two low-dose (Group 3) Main Study males were found dead (Days 151 and 198) and one middose (Group 5) Main Study male was sacrificed in a moribund condition on Day 55. All highdose Main Study males survived until terminal sacrifice. Two vehicle control (Group 1) Main Study females were sacrificed in a moribund condition (Days 165 and 178), two low-dose (Group 4) Main Study females were found dead (Days 119 and 197 [this death was determined to have been caused by a gavage error]), two mid-dose (Group 6) Main Study females were sacrificed in a moribund condition (Days 94 and 142) and one was found dead on Day 121. One high-dose (Group 8) Main Study female was found dead on Day 170. All other Main Study female mice survived until terminal sacrifice.

“Six male and four female positive control animals were either found dead or sacrificed in a moribund condition. The number of deaths in the male positive control animals was statistically significantly increased when compared to the vehicle control. All positive control animals were terminally sacrificed on study Days 121 (males) and 123 (females).

“In the TK animals, one Group 4 female was found dead on Day 17 and one Group 8 female was found dead on Day 153.” (page 27 of mice report)

Tumorigenicity analysis:

The Sponsor’s report indicates that tumors are analyzed using Peto methods where the logrank results on fatal, and mortality independent tumors were pooled with a life table analysis of incidental tumors. But it concludes that “There was no increase in the incidence of tumors in the test article treated mice of either sex.” (page 31).

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 table, Table 21 (a repeat of Table 4 and Table A.1.2 in Appendix 1), summarizes the results from tests comparing survival profiles across study groups in the tumorigenicity data sets:

Although of limited use, the Kaplan-Meier survival plots for mice are presented in figures A.1.3 and A.1.4 in Appendix 1. An alternative to the usual tests of survival is presented in the table below, displaying the actual observed times to death of the animals in each of the main study groups. Note the day of terminal sacrifice and the number of animals sacrificed are listed to the right under each gender.

Table 21. Survival times of Tg.rasH2 mice.

Dose group	Male Mice			Female Mice		
	Day of death	Term. Sac.		Day of death	Term. Sac.	
		Day	#		Day	#
1. Vehicle	97,142	197	23	78,165	197	23
2.Pos Ctrl	31, 78,107,107,116,118	121	9	106,118,119,119	123	11
3.Low	151,198	198	23	119,197	198	23
4.Medium	55	198	24	94,121,142	198	22
5.High	-	197	25	170	197	25

Except for the positive control there is little difference in death rates between study groups. In particular, it is apparent that in both genders there was little difference in survival between groups 1, 3, 4, and 5, and what little difference was manifest suggested a possible positive dose effect on survival. Statistical significance levels of exact logrank tests are also

provided in Appendix 1, but it was felt these tests were not needed for one to conclude that, except for the positive control, there was little dose effect on survival in either gender in mice.

Tumorigenicity analysis:

The table below displays the organ-tumor combinations that are statistically significant using the Haseman-Lin-Rahman (HLR) rules for adjusting for multiplicity. Note that only comparisons to the positive control achieve this level. More complete tables of the results of tests of tumor incidence in the various organ-tumor combinations in mice are given in Tables A.3.6. and A.3.7 in Appendix 3.

Table 22. Potentially Statistically Significant Neoplasms in Mice

Sex/ Organ/ tumor	Incidence					Significance Level		
	Veh	Pos.	Low	Med	High	ptrend	phigh vsVeh	pmed vsVeh/ plow vsVeh
Male Mice								
Systemic								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	14.0	24.5	25.0	25.0			
Hemangioma/Hemangiosarcoma	2	14	3	3	1	.8197	.8976	.5408
						.5209	<0.0001	
Adj. # at Risk	23.5	14.0	24.5	25.0	25.0			
hemangiosarcoma	2	14	2	3	1	.7584	.8976	.5408
						.7121	<0.0001	
lungs with bronchi								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	15.0	24.5	24.0	25.0			
Alv.Bronch. Adenoma/Carcinoma	7	15	3	3	2	.9467	.9922	.9699
						.9699	<0.0001	
Adj. # at Risk	23.5	15.0	24.5	24.0	25.0			
alveolar bronchiolar adenoma	7	15	3	3	2	.9467	.9922	.9699
						.9699	<0.0001	
Adj. # at Risk	23.5	9.8	24.5	24.0	25.0			
alveolar bronchiolar carcinoma	0	9	0	1	0	.5104	.	.5106
						.	<0.0001	
Adj. # at Risk	23.5	5.2	24.5	24.0	25.0			
hemangiosarcoma	0	3	0	0	0	.	.	.
						.	.0031	
spleen								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	14.0	24.5	24.0	25.0			
hemangiosarcoma	1	14	2	0	1	.6517	.7757	1
						.5163	<0.0001	
Female Mice								
Systemic								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	12.6	24.2	22.7	24.7			
hemangiosarcoma	0	12	1	0	2	.1134	.2553	.
						.5106	<0.0001	

Table 22. (cont.) Potentially Statistically Significant Neoplasms in Mice

Sex/ Organ/ tumor	Incidence					Significance Level		
	Veh	Pos. Ctrl	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh p _{vehvs} PosCtrl
lungs with bronchi								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	15.0	24.2	23.3	24.7			
Alv.Bronch. Adenocarcinoma	1	15	1	4	3	.1751	.3206	.1731
						.7660	<0.0001	
Adj. # at Risk	23.7	15.0	24.2	22.7	24.7			
alveolar bronchiolar adenoma	1	15	0	2	2	.1721	.5163	.4829
						1	<0.0001	
Adj. # at Risk	23.7	15.0	24.2	23.3	24.7			
alveolar bronchiolar carcinoma	0	15	1	2	1	.3302	.5106	.2444
						.5106	<0.0001	
spleen								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	12.6	24.2	22.7	24.7			
hemangiosarcoma	0	12	0	0	1	.2581	.5106	.
						.	<0.0001	

Thus, in both male and female mice, the only statistically significant results were the pairwise comparisons between the positive control and sterile water vehicle. In particular, in both genders in mice the differences between the positive control and vehicle were statistically significant in hemangiosarcomas of the spleen, and systemically (actually the same animals) and in alveolar adenoma, carcinoma, and adenocarcinoma or pooled adenoma/carcinoma.

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 Recommendation

Please see Section 1.1 above.

APPENDICES**Appendix 1. FDA Survival Analysis**

Simple summary life tables in mortality are presented in the report (Tables 16, 17, and 20 above). Kaplan-Meier estimated survival curves across study groups for each gender are displayed below in Figures A.1.1 and A.1.2 for rats and Figures A.1.3 and A.1.4 for mice. The plots include 95% confidence intervals around each survival curve (colored area around each curve), although the plots for mice are of arguable utility. These plots are also supported by tests of homogeneity in survival over the treatment groups. The statistical significance levels (i.e., p-values) are provided in Tables A.1.1. and A.1.3., below. One might note that the log rank tests place greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus, in the rat test, places more weight on differences in earlier events than does the log rank test. The tests in mice are exact test versions of the simple log rank test.

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over groups 1-5	0.0004	0.0003	0.1194	0.1684
Homogeneity over groups 1+2, 3-5	0.0002	0.0002	0.0654	0.0979
No Trend over dose groups 1+2, 3-5	< 0.0001	< 0.0001	0.1704	0.2480
No difference between groups 1+2 vs 5	< 0.0001	< 0.0001	0.1175	0.2080
No difference between groups 1 & 2	0.3645	0.1993	0.7820	0.7317

From Figure A.1.1 below, it seems that in male rats the high dose group tends to have the highest survival (i.e., lowest mortality) and the vehicle control group 1 generally with the the lowest survival. Through most of the study the remaining study groups are more or less intertwined until near the end of the study, when the mortality in the vehicle control group 2 increases to nearly match that in vehicle control group 1. As discussed in Section 1.3.1.1 below, this reviewer would argue that these apparent differences between the two essentially equivalent controls is likely due to the sort of random fluctuatiuons that occur in any study and should be ignored. However the differences in survival (in a negative direction) between the pooled vehicle group and the actual Cariprazine dose groups are sufficient to result in highly statistically significant tests of overall differences in survival among the three study groups and the pooled vehicle control (both logrank and Wilcoxon $p = 0.0002$). Similarly, tests of trend and pairwise differences between the high dose and vehicle were highly statistically significant (all four logrank and Wilcoxon $p < 0.0001$), though in a negative direction.

Figure A.1.2 is even simpler to interpret, with vehicle control groups 1 & 2 having the lowest survival, and the other study groups largely intertwined. These differences are sufficient to result in tests of lack of homogeneity of the three treatment groups and pooled control that are close to statistical significance at the usual 0.05 level (logrank $p = 0.0654$, Wilcoxon $p =$

0.0979), although again in a negative direction. No other test of trend or treatment group differences were statistically significant (all other eight $p \geq 0.1175$).

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

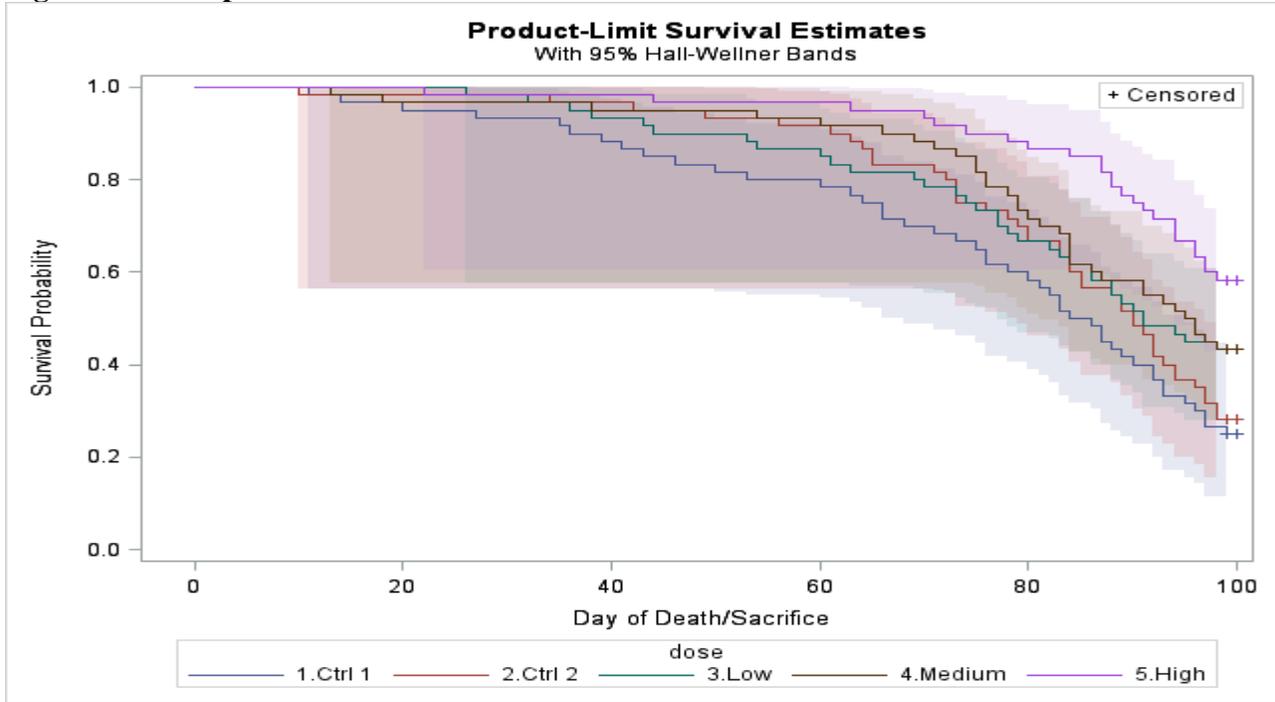


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats

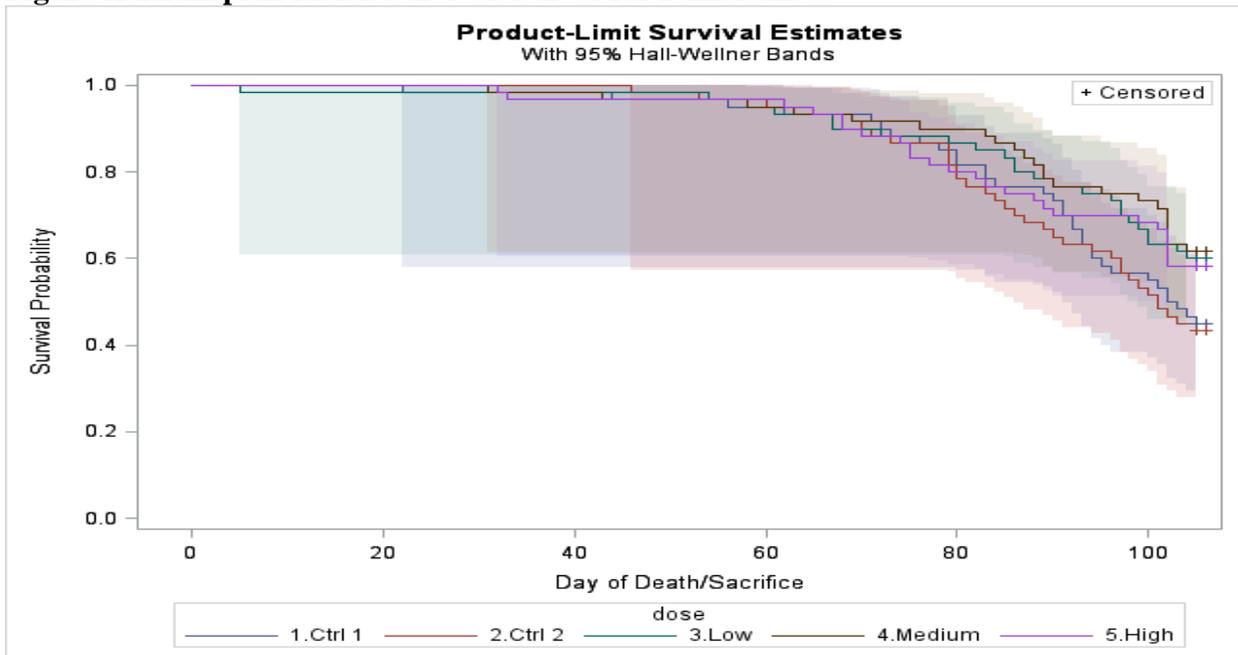


Table A.1.2, below, displays the actual observed times to death of the animals in each of the main study groups in the transgenic mouse study. Note the day of terminal sacrifice and the number of animals sacrificed are listed to the right under each gender.

Table A.1.2 Survival times of Tg.rasH2 mice.

Dose group	Male Mice			Female Mice		
	Day of death	Term. Sac.		Day of death	Term. Sac.	
		Day	#		Day	#
1. Vehicle	97,142	197	23	78,165	197	23
2.Pos Ctrl	31, 78,107,107,116,118	121	9	106,118,119,119	123	11
3.Low	151,198	198	23	119,197	198	23
4.Medium	55	198	24	94,121,142	198	22
5.High	-	197	25	170	197	25

The significance levels in the rat study above are based on large sample asymptotics. Due to the small number of events in the mouse study it seemed that the corresponding exact logrank tests would be more appropriate. Significance levels of these corresponding tests are based on Monte Carlo estimates of the permutation distribution of the test statistic. Results are presented below:

Table A.1.3 Statistical Significances of Exact Logrank Tests of Homogeneity in the Mouse Study

Hypothesis Tested	Males	Females
Homogeneity over groups 1-5	0.0001	0.0943
Homogeneity over groups 1, 3-5	0.7365	0.6721
No difference between groups 1 & 5	0.4898	0.4898
No difference between groups 1 & 2	0.0044	0.0417

From Table A.1.2 it is apparent that the positive control has a much greater number of deaths than the other dose groups. This is consistent with the results of the tests of homogeneity above. That is, in male mice the number of deaths in the other four dose groups is less than the number in the positive control. This is sufficient to result in a statistically significant test rejecting homogeneity over all five groups ($p = 0.0001$), as well as a pairwise test between vehicle and positive control ($p = 0.0044$). The somewhat similar pattern in female mice gives much weaker results. That is the tests of homogeneity are not as statistically significant ($p = 0.0943, 0.0417$, respectively).

Although of questionable utility, Figures A.1.3 and A.1.4, below, display the gender specific Kaplan-Meier survival curves over the five dose groups in mice.

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

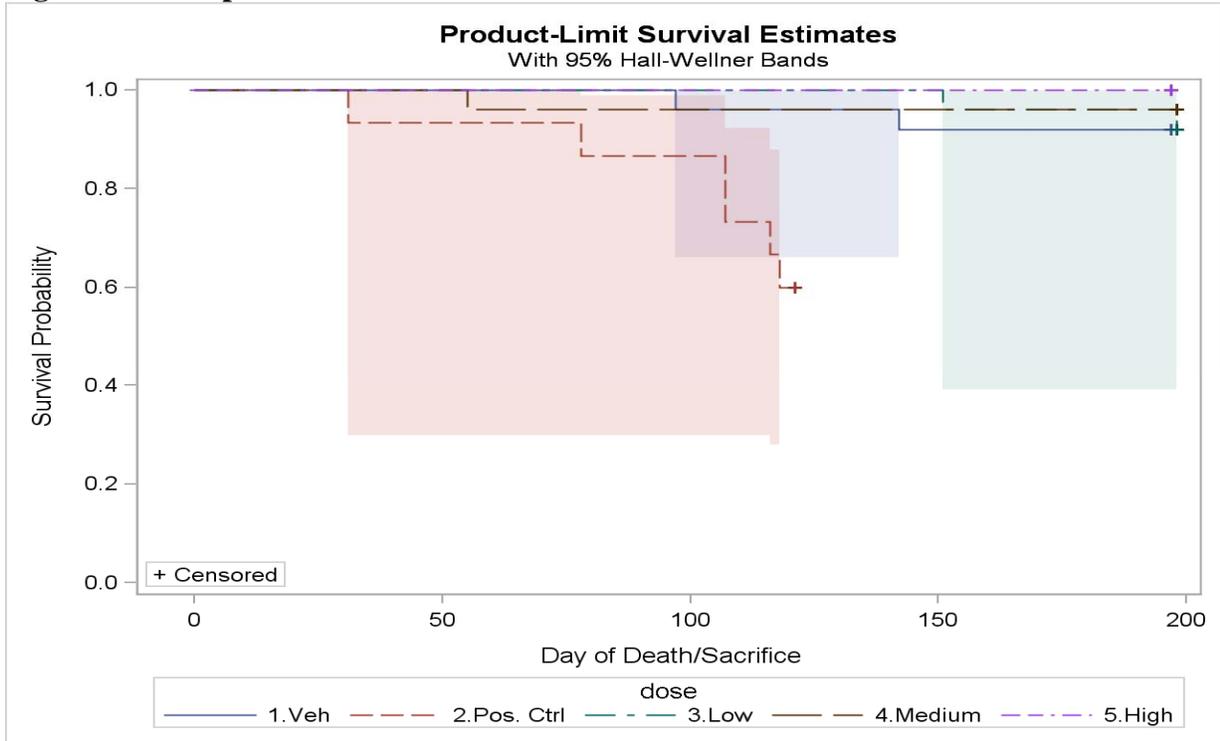
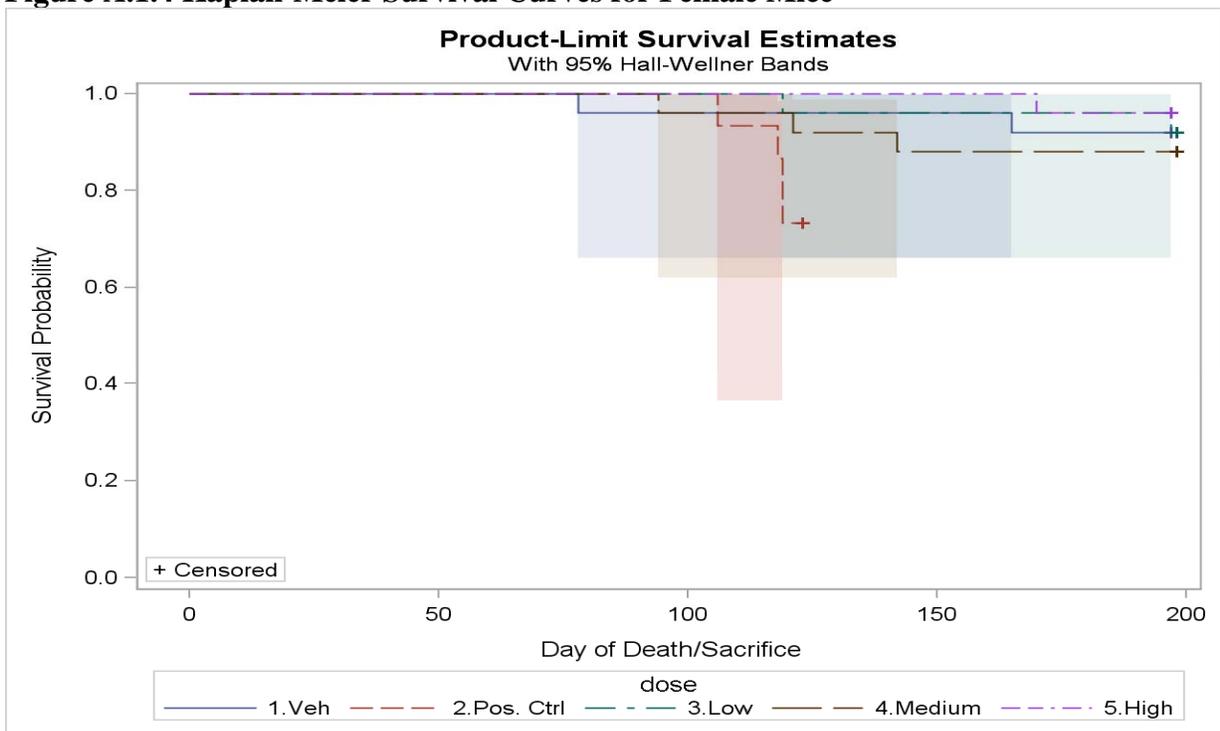


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



Appendix 2. FDA Nonparametric Bayesian Survival Analysis

The probability of a subject surviving past time t is given by the survival function, i.e., for random survival time T , $S(t) = P(T > t)$. Statistical inference on survival is based on proposing a probability model for $S(t)$ or one of its derivations. The probability model is defined so that hypotheses to be investigated are specified as parameters in the model. A frequentist analysis takes parameters as fixed and assesses the likelihood of the observed data. A Bayesian analysis starts by noting that parameters are not known, and assumes that a so-called prior probability distribution is a natural measure of this lack of exact knowledge about the parameters. Then the Bayesian analysis assesses the impact of the actual observed data on this prior. In a nonparametric Bayesian analysis at least one of these parameters is the space of all probability distributions, or at least some large subset of this space. In other words, although some prior weight is placed on a particular parametric family of distributions, the results would be consistent for other distributions and thus be robust to assumptions. The actual nonparametric analysis used here is based upon using a so-called Mixture of Dirichlet Processes (MDP) as the prior on the space of all probability distributions.

Specifically, let T_i denote a random variable representing the survival time of the i th animal. For time until natural death time t_i we write $T_i = t_i$, but if the animal is sacrificed at time a_i , all we know is that the time until natural death is greater than a_i , written as $T_i \in (a_i, \infty)$, i.e. T_i is in the time interval (a_i, ∞) . Note that animals whose death is in this interval are said to be censored. One useful probability model is to model the logarithm of T_i with a normal distribution, i.e., the T_i are modeled using a lognormal distribution. The mean of $\log T_i$ can be expressed as a product of a linear effect $X_i\beta$ times a usual lognormal term. Thus the linear effect accelerated (or decelerates) survival, justifying the Accelerated Failure Time (AFT) label for such a model. In this particular analysis we restrict attention to treated groups 2-5, assessing the effect of each of the three actual dose groups and the simple linear effect of dose over groups 3-5 where the baseline intercept is the vehicle effect. The distribution of $\log(T_i)$ is expressed as a mixture of normal distributions weighted by a Dirichlet process on the baseline normal parameters. Mathematically, we can write:

$$T_i = \exp(-X_i\beta)V_i, i = 1, \dots, n$$

$$V_i | G \sim G$$

$$\beta | \beta_0, S_{\beta_0} \sim N(\beta_0, S_{\beta_0})$$

$$G | \alpha, G_0 \sim DP(\alpha G_0)$$

The distributions of the hyperparameters above are specified as follows:

$$G_0 \sim \text{Lognormal}(\cdot | \mu, \sigma)$$

$$\alpha | a_0, b_0 \sim \text{Gamma}(a_0, b_0)$$

$$\mu | m_0, s_0 \sim N(m_0, s_0)$$

$$\sigma^{-1} | \tau_1, \tau_2 \sim \text{Gamma}\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right)$$

This analysis uses the DPsurvint function, for a Mixture of Dirichlet Processes in the DPpackage (Jara, 2007) of R (R Development Core Team, 2009). Currently, results should primarily be considered as supporting. The basic reference is de Iorio, et al (2009). The output for male rats follows:

**Table A.2.1 Output for Pairwise Differences From Vehicle in Male Rats
Bayesian Semiparametric AFT Regression Model**

Regression coefficients:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
d2	3.277e-02	3.280e-02	1.124e-03	3.026e-02	3.504e-02
d3	-2.335e-02	-2.347e-02	8.504e-04	-2.487e-02	-2.158e-02
d4	-1.193e-01	-1.194e-01	1.263e-03	-1.215e-01	-1.162e-01

Baseline distribution:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
mu	4.5120109	4.5115852	0.0402847	4.4343703	4.5933461
sigma2	0.3874995	0.3806386	0.0663551	0.2661699	0.5185779

Precision parameter:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
ncluster	161.17572	161.00000	13.21742	135.00000	186.00000
alpha	145.35731	140.70255	35.85321	82.67536	217.63313

Acceptance Rate for Metropolis Step = 0.4974023

The effects for d2, d3, and d4 represent the differences between the low, medium, and high dose groups, respectively, and the pooled vehicle group, where the latter is confounded with the baseline effect. For male rats, the 95% credible intervals for the effect of the high dose over the simple effect of the vehicle is about (-0.1215, -0.1162). Again, the posterior probability that the difference parameter is within those limits is about 0.95. Note that 0 is not in the credible interval, providing rather strong evidence the parameter is less than 0, corresponding to a decrease in deaths, i.e. an increase in survival. The credible interval for effect of the difference between the medium dose group and vehicle is (-0.0249,-0.0216), corresponding to an even greater increase in survival. The credible interval for the difference between the low dose and vehicle is (0.030, 0.035), this time suggesting an increase in death rate for the low dose over the vehicle.

Estimates are computed using Monte Carlo techniques on a Markov chain. The objective is to generate a rich pattern of feasible values for the parameters being analyzed. Proposed values are assessed if they fit the presumed model. If so, they are said to be accepted. The problem is that too high an acceptance rate is usually associated with small changes in the

proposed parameter values and thus induces high autocorrelations and poor searching over the space of possible values. For multivariate normal models an acceptance rate of somewhere between 0.2 to 0.25 is optimal, and, in general, the high acceptance rate above could be an indication that the estimated posterior distribution model may not be well estimated. However in this case, this does not seem to be true.

The trace is a plot of the posterior parameter estimate versus the iteration number. If the process is stationary, parameters can be estimates will be stable. In particular one looks for a flat, furry worm trace and a unimodal density for the posterior distribution of the parameter.

Figure A.2.1 Assessing Output for Dose Group Differences in Male Rats

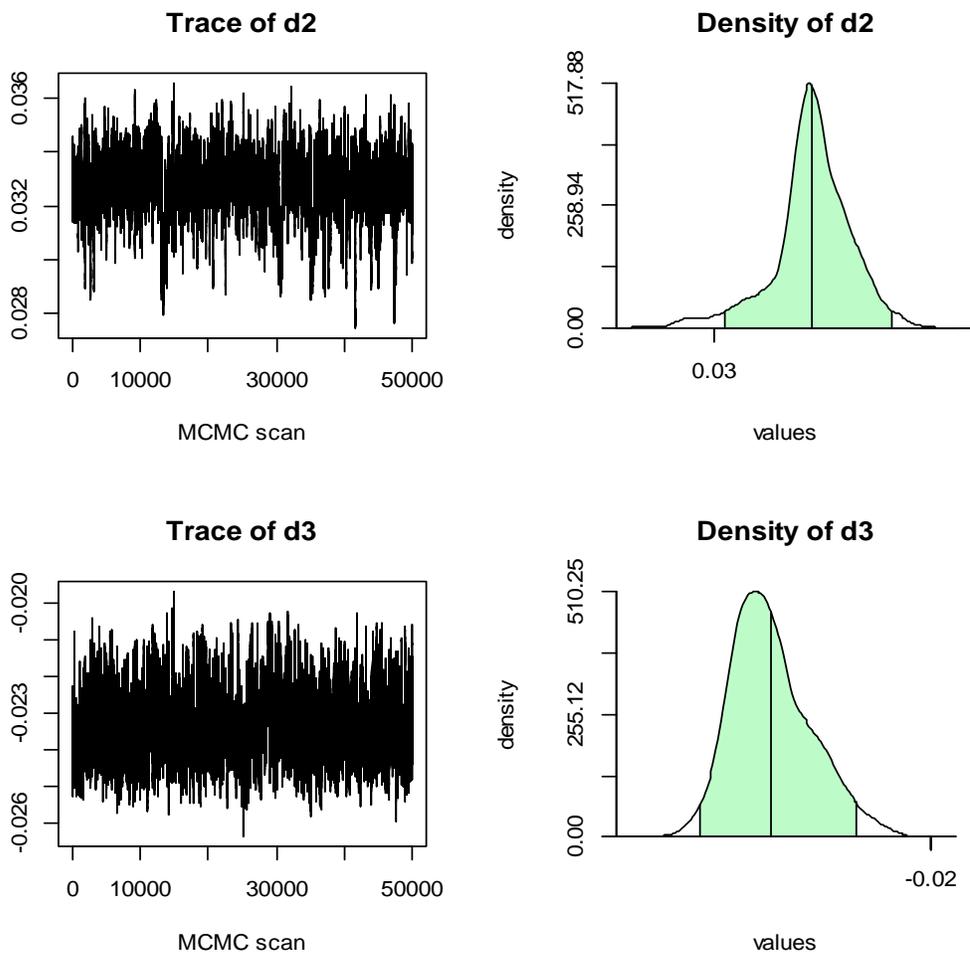
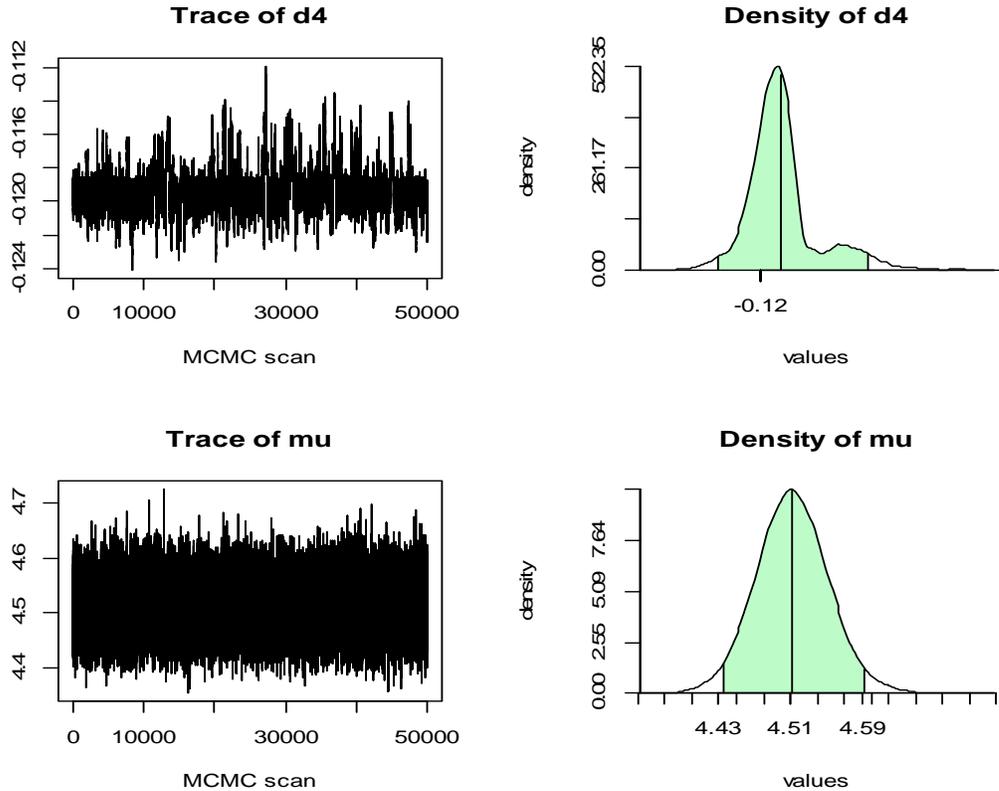


Figure A.2.1 (cont.) Assessing Output for Dose Group Differences in Male Rats



Several other assessments of convergence are needed, but clearly here the MCMC seems to be doing a reasonably good job in searching over the possible values of the parameters, and thus results should be dependable.

The following analysis attempt to addresses the slope parameter over the treatment groups the 1+2, 3-5 (i.e. pooled vehicle to high dose).

**Table A.2.2 Output for Dose Response Slope in Male Rats
Bayesian Semiparametric AFT Regression Model**

Regression coefficients:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
dose	-1.334e-02	-1.334e-02	1.513e-04	-1.364e-02	-1.305e-02

Baseline distribution:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
mu	4.5249681	4.5245573	0.0361091	4.4560517	4.5978466
sigma2	0.3742134	0.3679574	0.0597187	0.2682733	0.4959044

Table A.2.2 (cont.) Output for Dose Response Slope in Male Rats

Precision parameter:

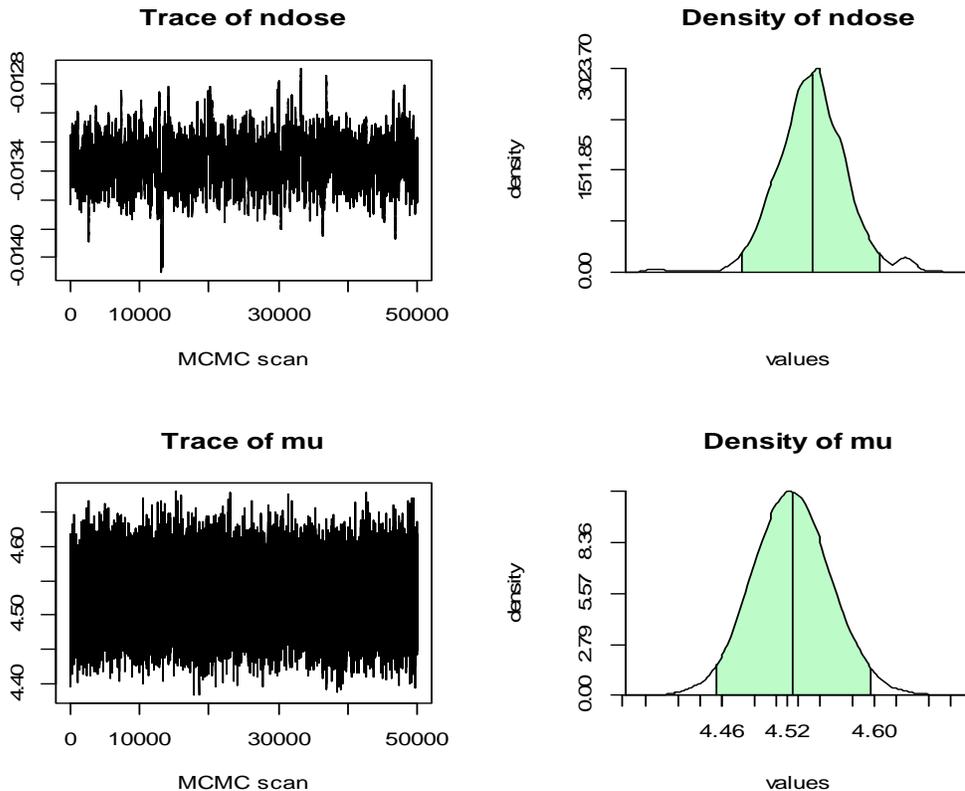
	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
ncluster	176.38578	176.00000	12.73248	152.00000	201.00000
alpha	184.22904	178.32393	44.95357	103.81347	273.31222

Acceptance Rate for Metropolis Step = 0.7147333

For male rats, the 95% credible intervals for the over all effect of dose in male rats in groups pooled vehicle through the high dose group is about (-0.0136,-0.0130). Usually the posterior probability the dose slope parameter is within those limits is about 0.95. Note that 0 is not in the interval. This provides rather strong evidence that the parameter is lesss than 0, corresponding to an increase in survival over increasing dose.

The trace plot of the posterior slope parameter estimate versus the interation number, given below, indicates that iterations do seem to be move through the space of feasible solutions, resulting in good ergodic behavior, and thus a good estimate.

Figure A.2.2 Assessing Output for Dose Response Slope in Male Rats



Note of course that the baseline mean, μ , is well estimated, but is probably of little interest.

Results for female rats are summarized below:

**Table A.2.3 Output for Pairwise Differences From Vehicle in Female Rats
Bayesian Semiparametric AFT Regression Model**

Regression coefficients:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
d2	8.854e-02	8.905e-02	4.766e-03	7.735e-02	9.478e-02
d3	1.050e-02	1.048e-02	8.838e-04	8.789e-03	1.225e-02
d4	6.154e-02	6.673e-02	1.052e-02	4.055e-02	7.419e-02

Baseline distribution:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
mu	4.8108611	4.8093665	0.0456249	4.7215260	4.9009378
sigma2	0.3494391	0.3423513	0.0629258	0.2371033	0.4763551

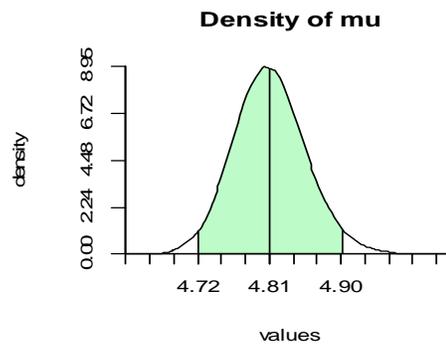
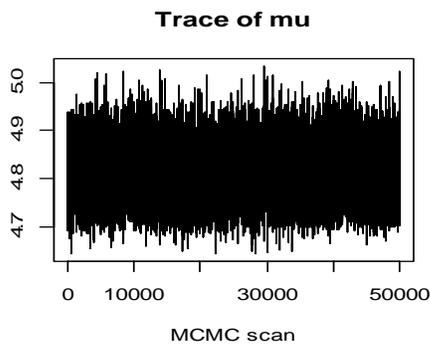
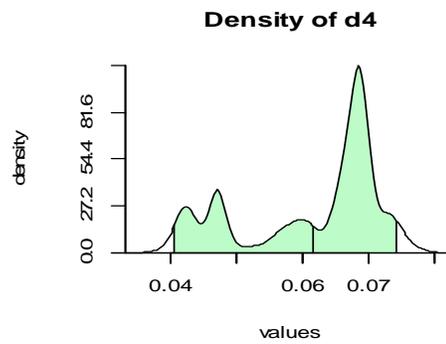
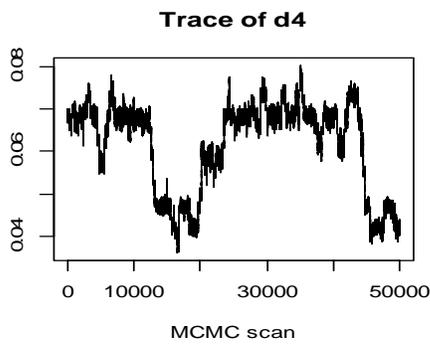
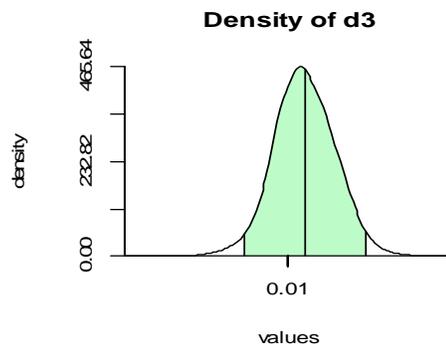
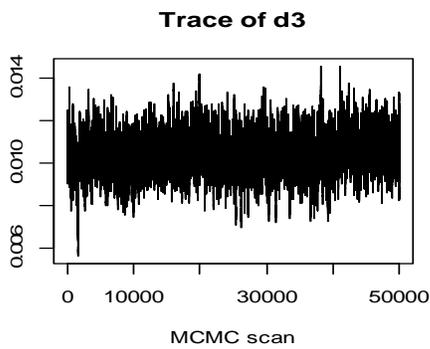
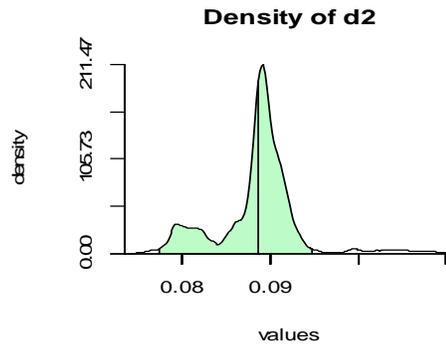
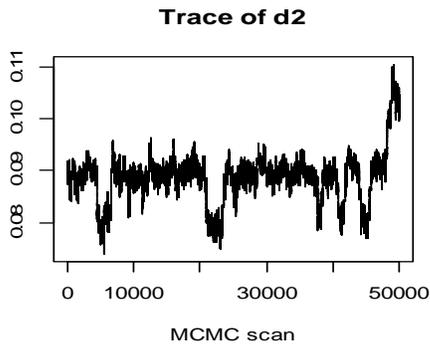
Precision parameter:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
ncluster	180.45340	180.00000	15.17680	150.00000	209.00000
alpha	198.67745	190.31224	56.61901	102.09850	309.52525

Acceptance Rate for Metropolis Step = 0.5948364

The trace plot below indicates that the d3, corresponding to the difference in effect between the medium dose and pooled vehicle control is well estimated and the probability is about 0.95 that the parameter is in the interval (0.0088, 0.0122), corresponding to a decrease in survival. The estimated parameters for d2 and d4 are also seem likely to be positive, but the trace plots and posterior distributions indicate they are not well estimated. Note that the trace plots for these posterior slope parameter estimate versus the iteration number, given below, indicates that iterations do seem to be largely stuck in separate regions of the space of feasible parameter values. This is also reflected in the estimated posterior densities. It may be due to small steps in the MCMC iterations, or poor fit to the AFT model, but this issue needs to be investigated.

Figure A.2.3 Assessing Output for Dose Group Differences in Female Rats



**Table A.2.4 Output for Dose Response Slope in Female Rats
Bayesian Semiparametric AFT Regression Model**

Regression coefficients:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
dose	5.075e-03	4.526e-03	8.485e-04	4.276e-03	6.621e-03

Baseline distribution:

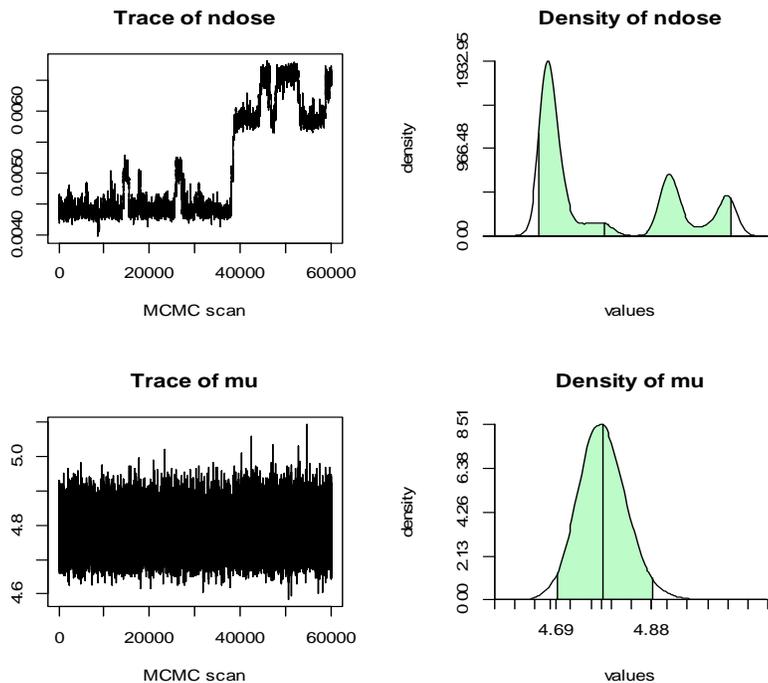
	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
mu	4.7830740	4.7814207	0.0476244	4.6939962	4.8815951
sigma2	0.3583213	0.3507395	0.0661181	0.2382237	0.4882561

Precision parameter:

	Mean	Median	Std. Dev.	95%HPD Lower	95%HPD Upper
ncluster	174.81137	175.00000	15.56193	144.00000	204.00000
alpha	181.98912	174.45417	52.19748	92.90387	286.45388

Acceptance Rate for Metropolis Step = 0.7517981

Figure A.2.4 Assessing Output for Dose Response Slope in Female Rats



Mixing for the estimation of the slope parameter is a real problem, with the statistical process supposedly iterating over the space of valid parameter values, and showing a major discontinuity after 38000-39000 iterations. This probably indicates that the AFT model is not appropriate for female rats. However, this issue requires further research and development. As in the other analyses, the overall mean does seem to be well estimated, but for this analysis, of little interest.

Appendix 3. FDA Poly-k Tumorigenicity Analysis

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. When there were no tumors of the specific type being analyzed in either column of the 2x2 table corresponding to a pairwise comparison an argument could be made that the p-value for this test should 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 a period “.”. Note that the StatXact program used for these analyses 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.

For each species by gender by organ the number of animals microscopically analyzed is presented first. Note that indicating an organ was not examined requires a specification in the data (please see section 2.2 above). This specification may be missing in some of this data. Thus, as discussed in Section 1.5 above, for some of these organs it is possibly more appropriate to define the actual endpoint used in the statistical analysis be the condition of being microscopically analyzed AND show the tumor. This does have problems unless treatment groups are not treated equally. The entry for each tumor is preceded by the adjusted number of animals at risk for that endpoint. It seems clear that an animal that dies early without having displaying that endpoint reduces the size of the risk set for that getting that particular endpoint. The poly-k test down weights such animals, and as discussed in Section 1.3.1.4, above, the sum of these poly-k weights seems to be a better estimate of the number of animals at risk of getting that tumor than the simple number of animals analyzed. This sum is given in the row labeled “Adjusted # at risk”. Tumor incidence is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high and medium dose groups versus vehicle. The next row continues with the p-values of the pairwise test between the low and vehicle dose groups and the p-values between the vehicle dose group and high dose group with water, respectively. For these analyses, incidence in the water only group is used to assess background tumor incidence, and thus whether a tumor is considered to be rare (background incidence <1%) or common. Note that for this analysis a tumor is only classified as rare if the H2O group shows none of that particular tumor.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman (HLR) rules discussed in Section 1.3.1.5 are often applied. That is, when testing for trend over dose groups 2-5 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 and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. Using these adjustments for other tests, like testing the comparisons between the low, medium, and water dose groups versus vehicle can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Tables A.3.1 and A.3.2 in rats and Tables A.3.3 and A.3.4 in mice show the tumors that had at least one mortality adjusted test whose nominal statistical significance was at least no more than 0.10. Note that when one adjusts for multiplicity these nominally significant comparisons may not be statistically significant.

Table A.3.1. Potentially Statistically Significant Neoplasms in Rats

Sex/ Organ/ Tumor	Incidence					Significance Level		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsveh
Male Rats								
PANCREAS								
# Evaluated	60	59	60	60	60			
Adj. # at Risk	35.4	38.7	39.6	43.4	50.1			
ISLET CELL ADENOMA	2	1	4	3	6	.0940	.0946	.3871
						.1850		
Adj. # at Risk	35.4	39.1	39.6	43.4	50.2			
Islet Cell Adenoma/Carcinoma	3	3	5	6	9	.0732	.0854	.2424
						.3122		
Female Rats								
ADRENAL GLANDS (PWG Analysis)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	46.6	45.9	49.4	50.8	48.9			
Medulla Pheochromocytoma [B&M]	1	3	1	1	8	.0020	.0178	.8902
						.8862		
ADRENAL GLANDS (Original Toxicologist)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.9	37.9	41.1	45.2	41.8			
MEDULLA BENIGN PHEOCHROMO- CYTOMA	1	1	0	1	4	.0226	.1119	.7559
						1		
Adj. # at Risk	38.9	37.9	41.1	45.2	42.8			
Medulla Pheochromocytoma [B&M]	1	1	0	1	5	.0072	.0541	.7559
						1		
PANCREAS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	38.4	42.1	45.6	41.7			
Islet Cell Adenoma/Carcinoma	0	3	5	2	2	.5700	.5700	.6109
						.1017		
SKIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.5	38.7	41.7	46.9	42.4			
FIBROMA	3	1	1	3	5	.0573	.1629	.5177
						.8844		
VAGINA								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	42.8			
SCHWANNOMA	0	0	0	0	2	.0416	.1247	.
						.		

Although all the organ tumor combinations in male rats listed above had at least one test that was statistically significant at a 0.10 level, adjusting for multiplicity, few were statistically significant. Using the HLR rules to adjust for multiplicity, note that the test of trend in pooled benign pheochromocytoma of the adrenals was statistically significant in the PWG analysis ($p = 0.002 < 0.005$) and close to significance using the tumor attribution used by the original toxicologist ($p = 0.0072 \approx 0.005$). No other comparisons achieved the multiplicity adjusted levels of statistical significance for a single study.

Table A.3.2. Potentially Statistically Significant Neoplasms in Mice

Sex/ Organ/ tumor	Incidence					Significance Level		
	Veh Ctrl	Pos.	Low	Med	High	ptrend vsVeh pVeh/ vsVeh	phigh vsVeh/ PosCtrl	pmed vsVeh/ vsVeh
Male Mice								
Systemic								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	14.0	24.5	25.0	25.0			
Hemangioma/Hemangiosarcoma	2	14	3	3	1	.8197 .5209	.8976 <0.0001	.5408
Adj. # at Risk	23.5	14.0	24.5	25.0	25.0			
hemangiosarcoma	2	14	2	3	1	.7584 .7121	.8976 <0.0001	.5408
lungs with bronchi								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	15.0	24.5	24.0	25.0			
Alv.Bronch. Adenoma/Carcinoma	7	15	3	3	2	.9467 .9699	.9922 <0.0001	.9699
Adj. # at Risk	23.5	15.0	24.5	24.0	25.0			
alveolar bronchiolar adenoma	7	15	3	3	2	.9467 .9699	.9922 <0.0001	.9699
Adj. # at Risk	23.5	9.8	24.5	24.0	25.0			
alveolar bronchiolar carcinoma	0	9	0	1	0	.5104 . <0.0001	. . <0.0001	.5106
Adj. # at Risk	23.5	5.2	24.5	24.0	25.0			
hemangiosarcoma	0	3	0	0	0	. . .0031	. . .0031	.
spleen								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	14.0	24.5	24.0	25.0			
hemangiosarcoma	1	14	2	0	1	.6517 .5163	.7757 <0.0001	1

Table A.3.2. (cont.) Potentially Statistically Significant Neoplasms in Mice

Sex/ Organ/ tumor	Incidence					Significance Level		
	Veh Ctrl	Pos.	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh PosCtrl
Female Mice								
Systemic								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	12.6	24.2	22.7	24.7			
hemangiosarcoma	0	12	1	0	2	.1134	.2553	.
						.5106	<0.0001	
lungs with bronchi								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	15.0	24.2	23.3	24.7			
Alv.Bronch. Adenocarcinoma	1	15	1	4	3	.1751	.3206	.1731
						.7660	<0.0001	
Adj. # at Risk	23.7	15.0	24.2	22.7	24.7			
alveolar bronchiolar adenoma	1	15	0	2	2	.1721	.5163	.4829
						1	<0.0001	
Adj. # at Risk	23.7	15.0	24.2	23.3	24.7			
alveolar bronchiolar carcinoma	0	15	1	2	1	.3302	.5106	.2444
						.5106	<0.0001	
spleen								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	12.6	24.2	22.7	24.7			
hemangiosarcoma	0	12	0	0	1	.2581	.5106	.
						.	<0.0001	

In both male and female mice, the only statistically significant results were the pairwise comparisons between the positive control and sterile water vehicle. In particular, in both genders in mice the differences between the positive control and vehicle were statistically significant in hemangiosarcomas of the spleen, and systemically (actually the same animals) and in alveolar adenoma, carcinoma, and adenocarcinoma or pooled adenoma/carcinoma.

Table A.3.3 displays the complete results of the Pathology Working Group (PWG) reanalysis of the rat adrenals. Tables A.3.4 and A.3.5 display all incidences and statistical test results for male and female rats, respectively, while Tables A.3.6 and A.3.7 present similar results in male and female mice. Again, 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.

Table A.3.3. PWG Identified Tumors in Male and Female Rats

Sex/ Organ/ Tumor	Incidence					Significance Level		
	Veh1	Veh2	Low	Med	High	ptrend	phigh vsVeh	pmed vsVeh/ plow vsveh
Male Rats								
ADRENAL GLANDS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.6	41.4	41.7	44.4	51.2			
CORTEX: ADENOMA	1	1	2	1	3	.2158	.3174	.7497
						.4382		
Adj. # at Risk	35.6	41.4	41.7	44.4	51.2			
Cortex Adenoma/Carcinoma	1	1	2	1	3	.2158	.3174	.7497
						.4382		
Adj. # at Risk	35.6	41.9	42.1	44.8	51.5			
MEDULLA BENIGN PHEOCHROMO- CYTOMA	2	4	4	5	6	.2511	.3238	.3635
						.4962		
Adj. # at Risk	35.6	41.0	41.5	44.9	51.2			
MEDULLA: GANGLIONEUROMA	0	0	0	1	0	.4481	.	.3667
						.		
Adj. # at Risk	35.6	41.0	41.5	44.4	51.7			
MEDULLA: MALIGNANT PHEOCHROMO- CYTOMA	0	1	0	0	2	.1444	.3531	1
						1		
Adj. # at Risk	35.6	41.9	42.1	44.8	52.0			
Medulla Pheochromocytoma [B&M]	2	5	4	5	7	.2012	.2937	.4564
						.5891		
Female Rats								
ADRENAL GLANDS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	46.6	45.2	49.4	50.9	47.9			
CORTEX: ADENOMA	4	1	0	2	1	.7398	.9224	.7808
						1		
Adj. # at Risk	46.3	45.2	49.4	50.7	48.0			
CORTEX: CARCINOMA	1	0	0	0	1	.3580	.5668	1
						1		
Adj. # at Risk	46.6	45.2	49.4	50.9	48.0			
Cortex Adenoma/Carcinoma	5	1	0	2	2	.5854	.8246	.8453
						1		
Adj. # at Risk	46.6	45.9	49.4	50.8	48.0			
MEDULLA BENIGN PHEOCHROMO- CYTOMA	1	3	1	1	6	.0164	.0739	.8902
						.8862		
Adj. # at Risk	46.3	45.2	49.4	50.7	48.9			
MEDULLA: MALIGNANT PHEOCHROMO- CYTOMA	0	0	0	0	2	.0400	.1176	.
						.		
Adj. # at Risk	46.6	45.9	49.4	50.8	48.9			
Medulla Pheochromocytoma [B&M]	1	3	1	1	8	.0020	.0178	.8902
						.8862		

Table A.3.4. Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend vsVeh	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
ADIPOSE TISSUE								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
LIPOMA	0	0	0	1	0	.4510	.	.3707
Adj. # at Risk	36.0	38.5	39.5	43.4	49.4	.		
MALIGNANT SCHWANNOMA	2	0	0	0	0	1	1	1
						1		
ADRENAL GLANDS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.9	39.7	43.4	49.4			
CORTEX: ADENOMA	1	1	2	1	3	.2128	.3117	.7508
						.4287		
Adj. # at Risk	35.1	38.9	39.7	43.4	49.4			
Cortex Adenoma/Carcinoma	1	1	2	1	3	.2128	.3117	.7508
						.4287		
Adj. # at Risk	35.1	39.4	40.1	43.7	50.2			
MEDULLA BENIGN PHEOCHROMO- CYTOMA	2	4	3	4	7	.1235	.2246	.5371
						.6734		
Adj. # at Risk	35.1	38.5	39.5	43.8	49.4			
MEDULLA: GANGLIONEUROMA	0	0	0	1	0	.4510	.	.3707
						.		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.9			
MEDULLA: MALIGNANT PHEOCHROMO- CYTOMA	0	1	0	0	2	.1440	.3532	1
						1		
Adj. # at Risk	35.1	39.4	40.1	43.7	50.6			
Medulla Pheochromocytoma [B&M]	2	5	3	4	8	.0967	.2065	.6302
						.7518		
BONE (OTHER)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.7			
AMYEOBLASTOMA	0	0	0	0	1	.2402	.4016	.
						.		
Adj. # at Risk	35.1	38.5	40.4	43.4	49.9			
OSTEOGENIC SARCOMA	0	0	1	0	1	.2507	.4016	.
						.3540		

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
BRAIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.7	49.5			
ASTROCYTOMA	0	0	0	1	1	.1586	.4016	.3707
						.		
Adj. # at Risk	35.7	38.7	39.5	43.6	49.4			
GRANULAR CELL TUMOR	1	1	0	2	0	.7745	1	.4688
						1		
Adj. # at Risk	35.1	38.5	40.0	43.4	49.4			
MENINGEAL SARCOMA	0	0	1	0	0	.6422	.	.
						.3482		
Adj. # at Risk	35.1	39.1	39.5	43.4	49.4			
OLIGODENDROGLIOMA	0	1	0	0	1	.4218	.6400	1
						1		
COLON								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.8	38.5	39.5	43.4	49.4			
ADENOCARCINOMA	1	0	0	0	0	1	1	1
						1		
DUODENUM								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
ADENOCARCINOMA	0	0	0	0	1	.2402	.4016	.
						.		
EAR(S)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.6	43.4	49.4			
FIBROSARCOMA	0	0	1	0	0	.6422	.	.
						.3482		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
SQUAMOUS CELL PAPILLOMA	0	0	1	0	0	.6422	.	.
						.3482		
EPIDIDYMIDES								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
MALIGNANT MESOTHELIOMA	0	1	0	0	0	1	1	1
						1		
HEAD								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.8	49.4			
OSTEOSARCOMA	0	0	0	1	0	.4510	.	.3707
						.		
Adj. # at Risk	35.1	38.5	40.4	43.4	49.4			
ZYMBAL'S GLAND CARCINOMA	0	0	1	0	0	.6439	.	.
						.3540		

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
HEART								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.7	38.5	39.5	43.4	49.4			
ENDOCARDIAL SCHWANNOMA	1	0	0	0	1	.4218	.6400	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
MYXOMA	0	0	1	0	0	.6422	.	.
						.3482		
JEJUNUM								
# Evaluated	60	60	59	60	60			
Adj. # at Risk	35.6	38.5	39.1	44.0	49.4			
ADENOCARCINOMA	1	1	0	3	0	.7838	1	.2681
						1		
KIDNEYS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	39.4	39.5	43.4	49.4			
NEPHROBLASTOMA	0	1	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
TUBULAR ADENOMA	1	0	0	0	1	.4236	.6440	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
TUBULAR CARCINOMA	0	1	0	0	0	1	1	1
						1		
LIVER								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
HEPATOCELLULAR ADENOMA	2	0	0	1	0	.8431	1	.7546
						1		
Adj. # at Risk	35.1	38.5	40.0	43.4	49.4			
HEPATOCELLULAR CARCINOMA	1	0	1	1	1	.4314	.6440	.6060
						.5772		
Adj. # at Risk	35.1	38.5	40.0	43.4	49.4			
Hepato. Adenoma/Carcinoma	3	0	1	2	1	.6940	.8761	.6145
						.8248		
LYMPH NODE OTHER								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.3	38.5	39.5	43.4	49.4			
HEMANGIOSARCOMA	1	0	0	0	0	1	1	1
						1		

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
LYMPH/RETIC SYS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	36.8	40.1	39.9	43.9	49.4			
HISTIOCYTIC SARCOMA	2	4	2	2	0	.9866	1	.8561
						.8240		
Adj. # at Risk	36.1	38.5	39.5	43.4	49.4			
LARGE GRANULAR LYMPHOCYTIC LUEKEMIA	1	0	0	0	0	1	1	1
						1		
Adj. # at Risk	36.0	39.0	40.4	43.9	49.4			
MALIGNANT LYMPHOMA	1	1	1	2	1	.5618	.7823	.4632
						.7265		
Adj. # at Risk	37.0	39.0	40.4	43.9	49.4			
Malig. Lymphoma/Lymph. Leuk.	2	1	1	2	1	.6791	.8676	.5968
						.8208		
MAMMARY AREAS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.3	38.5	39.5	43.4	49.5			
ADENOCARCINOMA	1	0	0	0	2	.1440	.3532	1
						1		
Adj. # at Risk	35.1	38.9	39.5	43.4	49.5			
ADENOMA	0	1	0	0	1	.4218	.6400	1
						1		
Adj. # at Risk	36.0	39.9	40.0	44.3	51.2			
Adenoma/-carcinoma/Fibro- sarcoma	2	3	2	2	6	.1113	.2481	.8064
						.7609		
Adj. # at Risk	35.8	39.5	40.0	44.3	51.0			
FIBROADENOMA	1	2	2	2	3	.3313	.4657	.6127
						.5595		
MEDIASTINAL LN								
# Evaluated	59	60	58	60	60			
Adj. # at Risk	35.1	38.5	39.3	43.7	49.4			
HEMANGIOMA	0	0	0	1	0	.4510	.	.3707
						.		
Adj. # at Risk	35.1	38.5	39.3	43.4	49.4			
HEMANGIOSARCOMA	0	0	0	0	1	.2402	.4016	.
						.		
MEDIASTINAL TISS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	44.2	49.4			
SCHWANNOMA	0	0	0	1	0	.4537	.	.3761
						.		

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
MESENTERIC LN								
# Evaluated	60	60	59	60	60			
Adj. # at Risk	35.3	38.9	38.6	43.4	49.4			
HEMANGIOMA	2	2	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	38.6	43.4	49.4			
HEMANGIOSARCOMA	0	0	0	0	1	.2414	.4016	.
						.		
MESENTERY/PERITO								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
FIBROSARCOMA	0	0	0	0	1	.2402	.4016	.
						.		
Adj. # at Risk	35.1	38.7	39.5	43.4	49.4			
HEMANGIOMA	0	2	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.4	50.0			
MYXOSARCOMA	0	0	0	0	1	.2439	.4065	.
						.		
MUSCLE (OTHER)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
HEMANGIOSARCOMA	0	0	1	0	0	.6422	.	.
						.3482		
PANCREAS								
# Evaluated	60	59	60	60	60			
Adj. # at Risk	35.1	38.3	39.5	43.4	49.4			
ACINAR CELL ADENOMA	0	3	0	0	1	.6785	.8761	1
						1		
Adj. # at Risk	35.1	38.1	39.6	43.7	49.4			
ACINAR CELL CARCINOMA	0	0	2	1	0	.6677	.	.3707
						.1192		
Adj. # at Risk	35.1	38.3	39.6	43.7	49.4			
Acinar cell Adenoma/Carc.	0	3	2	1	1	.7718	.8761	.8480
						.5716		
Adj. # at Risk	35.4	38.7	39.6	43.4	50.1			
ISLET CELL ADENOMA	2	1	4	3	6	.0940	.0946	.3871
						.1850		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.5			
ISLET CELL CARCINOMA	1	2	1	3	3	.2669	.4591	.3937
						.8248		
Adj. # at Risk	35.4	39.1	39.6	43.4	50.2			
Islet Cell Adenoma/Carc.	3	3	5	6	9	.0732	.0854	.2424
						.3122		

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
PARATHYROID								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
ADENOMA	0	0	1	0	0	.6422	.	.
								.3482
PITUITARY GLAND								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	40.9	44.1	42.1	47.0	51.0			
PARS DISTALIS-ADENOMA	21	24	17	19	23	.7164	.8561	.9409
								.9349
Adj. # at Risk	35.1	38.8	39.5	43.4	49.4			
PARS DISTALIS: CARCINOMA	0	1	0	0	0	1	1	1
								1
Adj. # at Risk	35.6	38.5	39.5	43.8	49.4			
PARS INTERMEDIA: ADENOMA	2	0	0	1	0	.8410	1	.7508
								1
Adj. # at Risk	40.9	44.4	42.1	47.0	51.0			
Pars Dist. Adenoma/Carcin.	21	25	17	19	23	.7508	.8841	.9546
								.9493
PREPUT/CLIT GL								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.9	49.4			
SQUAMOUS CELL PAPILOMA	0	0	0	1	0	.4510	.	.3707
								.
PROSTATE								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.6	38.5	40.0	43.4	49.8			
ADENOMA	1	0	2	0	2	.2762	.3484	1
								.2732
Adj. # at Risk	35.1	38.5	39.8	43.4	49.4			
HEMANGIOSARCOMA	0	0	1	0	0	.6422	.	.
								.3482
SALIVARY GLANDS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
FIBROMA	0	1	0	0	0	1	1	1
								1
Adj. # at Risk	35.1	38.9	39.5	43.4	49.4			
SARCOMA NOT OTHERWISE SPECIFIED	0	1	0	0	0	1	1	1
								1

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
SKIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	39.0	39.5	43.4	49.4			
BASOSQUAMOUS TUMOR	0	1	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
BENIGN TRICHOEPITHELI	0	1	0	0	0	1	1	1
						1		
Adj. # at Risk	35.8	40.0	40.7	43.4	50.8			
FIBROMA	5	5	3	0	4	.7737	.8899	1
						.8978		
Adj. # at Risk	35.8	39.4	40.8	43.7	50.4			
FIBROSARCOMA	1	1	2	1	1	.6598	.7875	.7470
						.4335		
Adj. # at Risk	35.2	38.5	39.5	43.4	49.4			
HAIR FOLLICLE/MATRIX TUMOR	1	0	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
HEMANGIOMA	0	0	0	0	1	.2402	.4016	.
						.		
Adj. # at Risk	35.1	39.0	40.5	44.1	49.4			
HEMANGIOSARCOMA	0	1	1	1	0	.7534	1	.6087
						.5807		
Adj. # at Risk	35.5	42.5	39.9	43.8	49.9			
KERATOACANTHOMA	2	8	3	2	2	.9540	.9813	.9693
						.8806		
Adj. # at Risk	35.9	38.5	40.2	43.4	49.5			
LIPOMA	3	1	2	0	3	.3904	.5801	1
						.6889		
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
MALIGNANT BASAL CELL TUMOR	0	1	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	39.5	43.7	49.4			
MALIGNANT SCHWANNOMA	0	0	0	1	0	.4510	.	.3707
						.		
Adj. # at Risk	35.1	38.5	39.5	43.9	49.4			
MYXOMA	0	0	0	1	0	.4510	.	.3707
						.		
Adj. # at Risk	36.1	38.7	39.5	43.4	49.4			
MYXOSARCOMA	1	1	0	0	0	1	1	1
						1		
Adj. # at Risk	35.1	38.5	39.7	43.4	49.4			
SEBACEOUS CELL CARCINOMA	0	0	1	0	0	.6422	.	.
						.3482		

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
SKIN (cont.)								
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
SQUAMOUS CELL CARCINOMA	0	0	0	0	1	.2402	.4016	.
Adj. # at Risk	35.1	39.2	39.5	43.9	49.4			
SQUAMOUS CELL PAPILLOMA	0	2	0	1	0	.8410	1	.7508
SOFT TISSUE								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.7	49.4			
FIBROMA	0	0	0	1	0	.4510	.	.3707
Adj. # at Risk	35.1	38.6	39.5	43.4	49.4			
MALIGNANT SCHWANNOMA	0	1	0	1	0	.6998	1	.6060
Systemic								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.3	39.2	39.5	43.7	49.4			
HEMANGIOMA	2	4	0	1	1	.8920	.9747	.9637
Adj. # at Risk	35.3	39.0	40.8	44.1	49.4			
HEMANGIOSARCOMA	1	1	3	1	2	.4836	.5234	.7571
Adj. # at Risk	35.3	39.7	40.8	44.4	49.4			
Hemangioma/Hemangiosarcoma	2	5	3	2	3	.7302	.8410	.9135
TAIL								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.6	49.4			
SQUAMOUS CELL PAPILLOMA	0	0	0	1	0	.4510	.	.3707
TESTES								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.7	38.6	39.5	44.8	49.8			
BENIGN INTERSTITIAL CELL TUMOR	2	2	0	4	4	.1638	.4000	.3399
Adj. # at Risk	35.1	38.9	39.6	43.4	49.4			
MALIGNANT MESOTHELIOMA	0	1	2	1	0	.8075	1	.6020
THORACIC SC								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	39.3	39.5	43.4	49.4			
MALIGNANT OLIGODENDROGLIOMA	0	1	0	0	0	1	1	1
Adj. # at Risk	35.1	38.5	39.5	44.0	49.4			
MIXED GLIOMA	0	0	0	1	0	.4510	.	.3707

Table A.3.4. (cont.) Neoplasms in Male Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
THYMUS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	35.1	38.5	39.5	43.4	49.4			
BENIGN THYMOMA	1	0	0	0	1	.4236	.6440	1
								1
THYROID								
# Evaluated	58	60	60	59	60			
Adj. # at Risk	35.9	38.5	39.8	43.7	49.6			
C-CELL ADENOMA	3	3	3	2	6	.1876	.3233	.8644
						.6597		
Adj. # at Risk	34.3	38.5	39.8	43.0	49.4			
C-CELL CARCINOMA	0	0	1	0	0	.6436	.	.
						.3514		
Adj. # at Risk	35.9	38.5	40.1	43.7	49.6			
C-cell Adenoma/Carcinoma	3	3	4	2	6	.2272	.3233	.8644
						.4905		
Adj. # at Risk	34.3	38.7	40.3	43.8	49.7			
FOLLICULAR CELL ADENOMA	0	1	3	1	2	.3657	.3532	.6060
						.1262		
Adj. # at Risk	34.6	38.9	39.5	43.0	49.4			
FOLLICULAR CELL CARCINOMA	1	1	1	1	0	.8727	1	.7481
						.7271		
Adj. # at Risk	34.6	39.2	40.3	43.8	49.7			
Foll.cell Adenoma/Carcin.	1	2	4	2	2	.6566	.6702	.6145
						.1997		

Table A.3.5. Neoplasms in Female Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
ADRENAL GLANDS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.9	38.6	41.1	45.2	41.7			
CORTEX: ADENOMA	4	4	1	2	2	.8037	.9209	.9394
						.9820		
Adj. # at Risk	38.6	37.9	41.1	45.1	41.8			
CORTEX: CARCINOMA	1	0	0	0	1	.3639	.5800	1
						1		
Adj. # at Risk	38.9	38.6	41.1	45.2	41.8			
Cortex Adenoma/Carcinoma	5	4	1	2	3	.6895	.8581	.9601
						.9887		
Adj. # at Risk	38.9	37.9	41.1	45.2	41.8			
MEDULLA BENIGN PHEOCHROMO- CYTOMA	1	1	0	1	4	.0226	.1119	.7559
						1		
Adj. # at Risk	38.6	37.9	41.1	45.1	42.7			
MEDULLA: MALIGNANT PHEOCHROMO- CYTOMA	0	0	0	0	1	.2059	.3559	.
						.		
Adj. # at Risk	38.9	37.9	41.1	45.2	42.8			
Medulla Pheochromocytoma [B&M]	1	1	0	1	5	.0072	.0541	.7559
						1		
BONE (OTHER)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.1	37.9	41.1	45.1	41.7			
OSTEOGENIC SARCOMA	1	0	0	0	0	1	1	1
						1		
BRAIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.2	41.7			
ASTROCYTOMA	0	0	0	1	0	.4236	.	.3719
						.		
Adj. # at Risk	38.6	37.9	41.6	45.1	41.7			
GRANULAR CELL TUMOR	1	0	2	0	0	.8638	1	1
						.2806		
Adj. # at Risk	40.1	37.9	41.1	45.1	41.7			
OLIGODENDROGLIOMA	2	0	0	0	0	1	1	1
						1		
COLON								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
FIBROSARCOMA	0	0	0	1	0	.4236	.	.3719
						.		
EAR(S)								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.9	37.9	41.1	45.1	41.7			
NEURAL CREST TUMOR	1	0	0	0	0	1	1	1
						1		

Table A.3.5. (cont.) Neoplasms in Female Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	phigh vsVeh	pmed vsVeh/ plow vsVeh
EYES								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
IRIS: MELANOMA	1	0	0	0	0	1	1	1
						1		
HARDERIAN GL								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.4	45.1	41.7			
FIBROSARCOMA	0	0	1	0	0	.6256	.	.
						.3504		
HEAD								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.5	41.7			
ZYMBAL'S GLAND CARCINOMA	0	0	0	1	0	.4236	.	.3719
						.		
HEART								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.4	37.9	41.1	45.1	41.7			
ENDOCARDIAL SCHWANNOMA	1	0	0	1	0	.6666	1	.6036
						1		
JEJUNUM								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.9	37.9	41.1	45.1	41.7			
ADENOCARCINOMA	1	0	0	0	0	1	1	1
						1		
Adj. # at Risk	39.0	37.9	41.1	45.1	41.7			
LEIOMYOMA	1	0	0	0	0	1	1	1
						1		
KIDNEYS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	42.5			
TRANSITIONAL CELL CARCINOMA	0	0	0	0	1	.2059	.3559	.
						.		
Adj. # at Risk	38.6	38.0	41.1	45.1	41.7			
TUBULAR ADENOMA	0	1	0	0	0	1	1	1
						1		
LIVER								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
CHOLANGIOCARCINOMA	0	0	0	1	0	.4236	.	.3719
						.		
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
HEPATOCELLULAR ADENOMA	0	1	0	1	1	.2912	.5800	.6074
						1		

Table A.3.5. (cont.) Neoplasms in Female Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
LYMPH/RETIC SYS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	38.0	41.1	45.1	41.9			
HISTIOCYTIC SARCOMA	0	1	0	0	2	.1045	.2806	1
Adj. # at Risk	39.2	37.9	41.1	45.1	41.7			
LARGE GRANULAR LYMPHOCYTIC LEUKEMIA	1	0	0	0	0	1	1	1
Adj. # at Risk	38.6	38.7	42.3	45.6	42.8			
MALIGNANT LYMPHOMA	1	2	3	1	3	.2792	.3568	.8459
Adj. # at Risk	39.2	38.7	42.3	45.6	42.8			
Malig.Lymphoma/Lymph.Leuk.	2	2	3	1	3	.3738	.4756	.9047
MAMMARY AREAS								
# Evaluated	60	59	60	60	60			
Adj. # at Risk	42.3	40.6	43.0	47.4	43.5			
ADENOCARCINOMA	8	6	4	6	3	.9290	.9726	.8157
Adj. # at Risk	39.0	37.5	43.6	45.1	43.4			
ADENOMA	1	0	3	1	3	.1229	.1337	.6074
Adj. # at Risk	51.4	48.7	54.9	53.5	50.2			
Adenoma/-carcinoma/Fibro.	31	29	32	26	22	.9766	.9788	.9290
Adj. # at Risk	48.5	46.7	51.6	51.3	48.5			
FIBROADENOMA	24	25	28	20	20	.9119	.9030	.9456
MESENTERIC LN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	38.0	41.1	45.1	41.7			
HEMANGIOMA	2	1	0	0	0	1	1	1
Adj. # at Risk	38.6	37.9	41.1	45.4	41.7			
HEMANGIOSARCOMA	0	0	0	1	0	.4236	.	.3719
MESENTERY/PERITO								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	38.3	41.1	45.1	41.7			
MALIGNANT MESOTHELIOMA	0	1	0	0	0	1	1	1

Table A.3.5. (cont.) Neoplasms in Female Rats

Organ/ Tumor	Incidence					Significance ptrend	Levels	
	Veh1	Veh2	Low	Med	High		phigh vsVeh	pmed vsVeh/ plow vsVeh
OVARIES								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
BENIGN THECAL CELL TUMOR	0	0	0	1	0	.4236	.	.3719
Adj. # at Risk	38.6	37.9	41.3	45.1	41.7			
HEMANGIOMA	0	0	1	0	0	.6256	.	.
Adj. # at Risk	38.6	38.2	41.8	45.1	41.7			
LUTEOMA	0	1	1	0	0	.8610	1	1
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
MALIGNANT GRANULOSA CELL TUMOR	0	1	0	0	0	1	1	1
Adj. # at Risk	38.6	37.9	41.1	45.5	41.7			
MALIGNANT SCHWANNOMA	0	0	0	1	0	.4236	.	.3719
Adj. # at Risk	38.9	37.9	41.1	45.1	41.7			
SERTOLI CELL TUMOR	1	0	0	0	0	1	1	1
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
TUBULOSTROMAL ADENOMA	0	0	0	0	1	.2020	.3504	.
PANCREAS								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
ACINAR CELL CARCINOMA	0	0	1	0	0	.6256	.	.
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
HEMANGIOSARCOMA	0	0	1	0	0	.6256	.	.
Adj. # at Risk	38.6	38.4	41.6	45.6	41.7			
ISLET CELL ADENOMA	0	2	3	2	2	.3699	.4382	.4766
Adj. # at Risk	38.6	37.9	41.6	45.1	41.7			
ISLET CELL CARCINOMA	0	1	2	0	0	.8638	1	1
Adj. # at Risk	38.6	38.4	42.1	45.6	41.7			
Islet Cell Adenoma/Carc.	0	3	5	2	2	.5700	.5700	.6109
PARATHYROID								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	42.4			
ADENOMA	0	0	0	0	1	.2059	.3559	.

Table A.3.5. (cont.) Neoplasms in Female Rats

Organ/ Tumor	Incidence					Significance ptrend	Levels phigh vsVeh	pmed vsVeh/ pmed vsVeh
	Veh1	Veh2	Low	Med	High			
PITUITARY GLAND								
# Evaluated	60	59	60	60	60			
Adj. # at Risk	51.2	49.5	50.4	52.5	51.5			
PARS DISTALIS-ADENOMA	37	34	31	37	33	.7134	.8355	.5701
						.9024		
Adj. # at Risk	39.5	37.6	41.1	45.1	42.1			
PARS DISTALIS: CARCINOMA	2	1	1	0	1	.6926	.8296	1
						.8237		
Adj. # at Risk	38.6	36.9	41.1	45.4	42.2			
PARS INTERMEDIA: ADENOMA	0	1	0	1	1	.2991	.5911	.6113
						1		
Adj. # at Risk	52.1	50.2	50.4	52.5	51.9			
Pars Dist.Adenoma/Carcin.	39	35	32	37	34	.7078	.8269	.6480
						.8965		
PREPUT/CLIT GL								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
SQUAMOUS CELL PAPILOMA	1	0	0	0	0	1	1	1
						1		
SKIN								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.5	38.7	41.7	46.9	42.4			
FIBROMA	3	1	1	3	5	.0573	.1629	.5177
						.8844		
Adj. # at Risk	39.2	38.4	41.1	45.5	42.5			
FIBROSARCOMA	1	1	0	1	1	.4548	.7329	.7522
						1		
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
HEMANGIOSARCOMA	1	0	0	0	0	1	1	1
						1		
Adj. # at Risk	39.1	37.9	41.1	45.1	41.7			
LEIOMYOSARCOMA	1	0	0	0	0	1	1	1
						1		
Adj. # at Risk	38.7	37.9	41.1	45.1	41.7			
LIPOMA	1	0	0	1	0	.6690	1	.6074
						1		
Adj. # at Risk	38.6	37.9	41.1	45.1	42.5			
MALIGNANT BASAL CELL TUMOR	0	0	0	0	1	.2059	.3559	.
						.		
Adj. # at Risk	38.6	38.4	41.1	45.2	41.7			
SQUAMOUS CELL CARCINOMA	0	1	1	1	0	.7141	1	.6036
						.5761		
SOFT TISSUE								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7			
SARCOMA NOT OTHERWISE SPECI- FIED	0	0	0	1	0	.4236	.	.3719
						.		

Table A.3.5. (cont.) Neoplasms in Female Rats

Organ/ Tumor	Incidence					ptrend	Significance Levels		
	Veh1	Veh2	Low	Med	High		phigh vsVeh	pmed vsVeh/	pplow vsVeh
SPLEEN									
# Evaluated	60	60	60	60	60				
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7				
HEMANGIOMA	0	0	0	0	1	.2020	.3504	.	.
Systemic									
# Evaluated	60	60	60	60	60				
Adj. # at Risk	38.6	38.0	41.3	45.1	41.7				
HEMANGIOMA	2	1	1	0	1	.6884	.8269	1	
Adj. # at Risk	38.6	37.9	41.1	45.4	41.7	.8269			
HEMANGIOSARCOMA	1	0	1	1	0	.7173	1	.6074	
Adj. # at Risk	38.6	38.0	41.3	45.4	41.7	.5800			
Hemangioma/Hemangiosarcoma	3	1	2	1	1	.7954	.8897	.9071	
Adj. # at Risk	38.6	38.0	41.3	45.4	41.7	.6879			
THYMUS									
# Evaluated	60	60	60	60	60				
Adj. # at Risk	38.6	37.9	41.8	45.1	41.7				
BENIGN THYMOMA	0	0	1	0	0	.6256	.	.	
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7	.3504			
MALIGNANT THYMOMA	0	1	0	0	0	1	1	1	
Adj. # at Risk	38.6	37.9	41.1	45.1	41.7	1			
THYROID									
# Evaluated	60	60	60	60	59				
Adj. # at Risk	40.1	39.3	42.7	45.2	43.7				
C-CELL ADENOMA	6	5	3	6	7	.2594	.4595	.6352	
Adj. # at Risk	38.6	37.9	41.7	45.1	41.1	.9255			
C-CELL CARCINOMA	0	0	2	0	0	.7090	.	.	
Adj. # at Risk	38.6	37.9	41.7	45.1	41.1	.1208			
C-cell Adenoma/Carcinoma	40.1	39.3	43.2	45.2	43.7	.3227	.4595	.6352	
Adj. # at Risk	38.9	38.5	41.4	45.1	41.1	.7340			
FOLLICULAR CELL ADENOMA	1	2	1	0	0	.9807	1	1	
Adj. # at Risk	38.6	38.4	41.6	45.1	41.1	.8237			
FOLLICULAR CELL CARCINOMA	0	1	1	0	0	.8587	1	1	
Adj. # at Risk	38.9	39.0	41.8	45.1	41.1	.5761			
Foll.cell Adenoma/Carcin.	1	3	2	0	0	.9887	1	1	
Adj. # at Risk	38.9	39.0	41.8	45.1	41.1	.6822			

Table A.3.5. (cont.) Neoplasms in Female Rats

Organ/ Tumor	Incidence					Significance Levels		
	Veh1	Veh2	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ p _{low} vsVeh
UTERUS W/ CERVIX								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	38.6	37.9	42.1	45.1	41.7			
DECIDUOMA	0	0	1	0	0	.6275	.	.
						.3559		
Adj. # at Risk	38.6	38.4	41.1	45.1	41.7			
ENDOMETRIAL CARCINOMA	0	1	0	0	0	1	1	1
						1		
Adj. # at Risk	39.5	37.9	41.1	45.5	41.7			
ENDOMETRIAL STROMAL POLYP	3	3	2	4	1	.8491	.9544	.5406
						.8365		
Adj. # at Risk	38.6	37.9	41.1	45.2	41.7			
ENDOMETRIAL STROMAL SARCOMA	0	0	0	1	1	.1300	.3504	.3719
						.		
Adj. # at Risk	39.5	37.9	41.1	45.7	41.7			
Endo.Stromal Polyp/Sarcoma	3	3	2	5	2	.6574	.8365	.3781
						.8365		
Adj. # at Risk	38.6	38.2	41.1	45.1	41.7			
LEIOMYOSARCOMA	0	1	0	0	0	1	1	1
						1		
Adj. # at Risk	39.1	38.6	41.1	45.1	41.7			
MALIGNANT SCHWANNOMA	1	1	0	0	1	.5020	.7259	1
						1		
VAGINA								
# Evaluated	60	60	60	60	60			
Adj. # at Risk	39.5	37.9	41.1	45.1	41.7			
BENIGN GRANULAR CELL TUMOR	2	0	0	0	0	1	1	1
						1		
Adj. # at Risk	39.3	37.9	41.1	45.1	41.7			
SARCOMA NOT OTHERWISE SPECI- FIED	1	0	0	0	0	1	1	1
						1		
Adj. # at Risk	38.6	37.9	41.1	45.1	42.8			
SCHWANNOMA	0	0	0	0	2	.0416	.1247	.
						.		

Table A.3.6. Neoplasms in Male Mice

Organ/ Tumor	Incidence					Significance Levels		
	Veh Ctrl	Pos.	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ vs PosCtrl
Systemic								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	14.0	24.5	25.0	25.0			
Hemangioma/Hemangiosarcoma	2	14	3	3	1	.8197 .5209	.8976 <0.0001	.5408
Adj. # at Risk hemangioma	23.5 0	2.9 0	24.5 1	24.0 0	25.0 0	.7604 .5106
Adj. # at Risk hemangiosarcoma	23.5 2	14.0 14	24.5 2	25.0 3	25.0 1	.7584 .7121	.8976 <0.0001	.5408
bone marrow, femur								
# Evaluated	25	0	25	25	25			
Adj. # at Risk hemangiosarcoma	23.5 0	0.0 0	24.5 0	24.0 1	25.0 0	.51045106
bone marrow, sternum								
# Evaluated	25	0	25	25	25			
Adj. # at Risk hemangiosarcoma	23.5 1	0.0 0	24.5 0	24.0 0	25.0 0	1 1	1 .	1
harderian glands								
# Evaluated	25	0	25	25	25			
Adj. # at Risk adenoma	23.5 0	0.0 0	24.5 0	24.0 2	25.0 0	.51562553
lungs with bronchi								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	15.0	24.5	24.0	25.0			
Alv.Bronch. Adenoa/Carcinoma	7	15	3	3	2	.9467 .9699	.9922 <0.0001	.9699
Adj. # at Risk alveolar bronchiolar adenoma	23.5 7	15.0 15	24.5 3	24.0 3	25.0 2	.9467 .9699	.9922 <0.0001	.9699
Adj. # at Risk alveolar bronchiolar carcinoma	23.5 0	9.8 9	24.5 0	24.0 1	25.0 0	.5104 .	. <0.0001	.5106
Adj. # at Risk hemangiosarcoma	23.5 0	5.2 3	24.5 0	24.0 0	25.0 00031	. .
mandible								
# Evaluated	1	0	0	0	0			
Adj. # at Risk squamous cell carcinoma	1.0 1	0.0 0	0.0 0	0.0 0	0.0 0	1

Table A.3.6. (cont.) Neoplasms in Male Mice

Organ/ Tumor	Incidence					Significance Levels		
	Veh Ctrl	Pos. Ctrl	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ vs PosCtrl
multicentric								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	2.9	24.5	25.0	25.0			
hemangiosarcoma	0	0	0	1	0	.5155	.	.5208
						.	.	
Adj. # at Risk	23.5	3.7	24.5	24.0	25.0			
lymphoma	0	1	0	0	0	.	.1154	.
						.		
Adj. # at Risk	23.5	2.9	25.0	24.0	25.0			
mesothelioma	0	0	1	0	0	.7629	.	.
						.5208	.	
salivary glands								
# Evaluated	25	0	25	25	25			
Adj. # at Risk	23.5	0.0	24.5	24.0	25.0			
hemangioma	0	0	1	0	0	.7604	.	.
						.5106	.	
spleen								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.5	14.0	24.5	24.0	25.0			
hemangiosarcoma	1	14	2	0	1	.6517	.7757	1
						.5163	<0.0001	
stomach								
# Evaluated	25	0	25	25	25			
Adj. # at Risk	23.5	0.0	24.5	24.0	25.0			
squamous cell carcinoma	0	0	0	1	0	.5104	.	.5106
						.	.	
testes								
# Evaluated	25	0	25	25	25			
Adj. # at Risk	23.5	0.0	24.5	24.0	25.0			
hemangiosarcoma	0	0	0	1	0	.5104	.	.5106
						.	.	
Adj. # at Risk	23.5	0.0	24.5	24.0	25.0			
interstitial cell adenoma	0	0	0	0	1	.2604	.5208	.
						.	.	
Systemic								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	12.6	24.2	22.7	24.7			
hemangiosarcoma	0	12	1	0	2	.1134	.2553	.
						.5106	<0.0001	

Table A.3.7. Neoplasms in Female Mice

Organ/ Tumor	Incidence					Significance Levels		
	Veh Ctrl	Pos.	Low	Med	High	ptrend	phigh vsVeh	pmed vsVeh/ PosCtrl
harderian glands								
# Evaluated	25	0	25	25	25			
Adj. # at Risk	23.7	0.0	24.2	22.7	24.7			
adenoma	1	0	0	0	0	1	1	1
						1	.	
lungs with bronchi								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	15.0	24.2	23.3	24.7			
Alv.Bronch. Adenoma/Carcinoma	1	15	1	4	3	.1751	.3206	.1731
						.7660	<0.0001	
Adj. # at Risk	23.7	15.0	24.2	22.7	24.7			
alveolar bronchiolar adenoma	1	15	0	2	2	.1721	.5163	.4829
						1	<0.0001	
Adj. # at Risk	23.7	15.0	24.2	23.3	24.7			
alveolar bronchiolar carcinoma	0	15	1	2	1	.3302	.5106	.2444
						.5106	<0.0001	
Adj. # at Risk	23.7	3.5	24.2	22.7	24.7			
hemangiosarcoma	0	0	0	0	1	.2581	.5106	.
						.	.	
multicentric								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	24.6	3.5	24.2	22.7	24.7			
lymphangioma	1	0	0	0	0	1	1	1
						1	1	
Adj. # at Risk	24.1	3.5	24.2	22.7	24.7			
mesothelioma	1	0	0	0	0	1	1	1
						1	1	
perineum								
# Evaluated	0	0	0	2	0			
Adj. # at Risk	0.0	0.0	0.0	2.0	0.0			
papilloma	0	0	0	1	0	1	.	.
						.	.	
Adj. # at Risk	0.0	0.0	0.0	1.1	0.0			
squamous cell carcinoma	0	0	0	1	0	1	.	.
						.	.	
spleen								
# Evaluated	25	15	25	25	25			
Adj. # at Risk	23.7	12.6	24.2	22.7	24.7			
hemangiosarcoma	0	12	0	0	1	.2581	.5106	.
						.	<0.0001	
Adj. # at Risk	23.7	4.3	24.2	22.7	24.7			
histiocytic sarcoma	0	1	0	0	0	.	.	.
						.	.1481	

Table A.3.7. (cont.) Neoplasms in Female Mice

Organ/ Tumor	Incidence					Significance Levels		
	Veh Ctrl	Pos. Ctrl	Low	Med	High	ptrend	p _{high} vsVeh	p _{med} vsVeh/ vs PosCtrl
stomach								
# Evaluated	25	0	25	25	25			
Adj. # at Risk	23.7	0.0	24.2	22.7	24.7			
papilloma	0	0	0	0	1	.2581	.5106	.
						.	.	
Adj. # at Risk	23.7	0.0	24.2	22.7	24.7			
squamous cell carcinoma	0	0	0	1	0	.4946	.	.4889
						.	.	
uterus								
# Evaluated	25	0	25	25	25			
Adj. # at Risk	23.7	0.0	24.2	22.7	24.7			
hemangiosarcoma	0	0	1	0	0	.7527	.	.
						.5106	.	

Appendix 4. 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.
- Greaves, P. (2007), “Neoplasia of Adrenal Medulla,” In: *Histopathology of Preclinical Toxicity Studies*. 3rd edition, pp 819. Oxford, UK: Academic Press, Elsevier Inc.
- Haseman, J. K. (1983), “A Reexamination of False-positive Rates for Carcinogenicity Studies”, *Fundamental and Applied Toxicology*, **3**, 334-339.
- Jara, A. (2007), “Applied Bayesian Non- and Semi-parametric Inference using DPpackage”, *Rnews*, **7**, 3, 17-26.
- 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.
- Parola, A, and Jacobs, A. (2010). “Combining Tumors for Statistical Analysis”, online FDA handout.
- 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.

NDA 204370 Cariprazine

Forest Research Institute

R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

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 (2013), 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

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
07/11/2013
Statistical Carcinogenicity Review

KARL K LIN
07/11/2013
Concur with review