

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

202611Orig1s000

STATISTICAL REVIEW(S)



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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA Serial Number: 202611

Drug Name: Mirabegron

Indication(s): Overactive bladder

Applicant: Astellas Pharma Global Development, INC.

Date(s): Submission date: 08/29/2011
PDUFA date: 06/29/2012

Review Priority: Standard

Biometrics Division: Division of Biometrics III

Statistical Reviewer: Jia Guo, Ph.D.

Concurring Reviewers: Mahboob Sobhan, Ph.D.

Medical Division: Division of Reproductive and Urologic Products, HFD-580

Clinical Team: Rodger Wiederhorn, M.D., Clinical Reviewer
Mark Hirsch, M.D., Clinical Team Leader

Project Manager: Nenita Crisostomo

Keywords: analysis of covariance, multiple comparisons, multiple endpoints

Table of Contents

1	EXECUTIVE SUMMARY	5
2	INTRODUCTION	7
2.1	OVERVIEW	7
2.2	DATA SOURCES	7
3	STATISTICAL EVALUATION	8
3.1	DATA AND ANALYSIS QUALITY	8
3.2	EVALUATION OF EFFICACY	8
3.2.1	Study Design and Endpoints.....	8
3.2.1.1	Study Design	8
3.2.1.2	Primary/Key Secondary Efficacy Endpoints	8
3.2.1.3	Multiplicity Control Approach	9
3.2.2	Patient Disposition, Demographic and Baseline Characteristics	10
3.2.2.1	Patient Disposition	10
3.2.2.2	Demographics and Baseline Characteristics.....	12
3.2.3	Statistical Methodologies	12
3.2.3.1	Applicant’s analyses.....	12
3.2.3.2	Reviewer’s analyses	13
3.2.4	Results and Conclusions	14
3.3	EVALUATION OF SAFETY	19
4	FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	19
4.1	GENDER, RACE, AGE, AND GEOGRAPHIC REGION.....	19
4.1.1	Gender	20
4.1.2	Age group.....	20
4.1.3	Race.....	20
4.1.4	Region	20
4.2	OTHER SPECIAL/SUBGROUP POPULATIONS.....	20
5	SUMMARY AND CONCLUSIONS	21
5.1	STATISTICAL ISSUES AND COLLECTIVE EVIDENCE.....	21
5.2	CONCLUSIONS AND RECOMMENDATIONS	21
	APPENDICES	22

LIST OF TABLES

Table 1: List of all Studies included in the Statistical Review	7
Table 2: Summary of Subject Disposition – Study 178-CL-046	10
Table 3: Summary of Subject Disposition – Study 178-CL-047	11
Table 4: Summary of Subject Disposition – Study 178-CL-074	11
Table 5: Summary of Efficacy analysis sets – Study 178-CL-046	11
Table 6: Summary of Efficacy analysis sets – Study 178-CL-047	12
Table 7: Summary of Efficacy analysis sets – Study 178-CL-074	12
Table 8: Analysis Approach for the Primary/Key Secondary Efficacy endpoints	13
Table 9: Summary of Primary and Key Secondary Efficacy endpoints – Study 178-CL-046	14
Table 10: Summary of Primary and Key Secondary Efficacy endpoints – Study 178-CL-047	15
Table 11: Summary of Primary and Key Secondary Efficacy endpoints – Study 178-CL-074	16
Table 12: Subgroup categories defined in each study	19
Table 13 Summary of Patient Demographics and Baseline Characteristics - Study 178-CL-046, SAF	23
Table 14 Summary of Patient Demographics and Baseline Characteristics - Study 178-CL-047, SAF	24
Table 15 Summary of Patient Demographics and Baseline Characteristics - Study 178-CL-074, SAF	25
Table 16: Summary of Additional Key Secondary Efficacy endpoints - Study 178-CL-074	26
Table 17: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Gender – Study 178-CL-046	27
Table 18: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Gender – Study 178-CL-047	27
Table 19: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Gender – Study 178-CL-074	27
Table 20: Change from Baseline to Final visit in Micturitions episodes per 24 hours by Gender – Study 178-CL-046	27
Table 21: Change from Baseline to Final visit in Micturitions episodes per 24 hours by Gender – Study 178-CL-047	28
Table 22: Change from Baseline to Final visit in Micturitions episodes per 24 hours by Gender – Study 178-CL-074	28
Table 23: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Age group – Study 178-CL-046	28
Table 24: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Age group – Study 178-CL-047	28
Table 25: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Age group – Study 178-CL-074	28
Table 26: Change from Baseline to Final visit in Micturitions per 24 hours by Age group – Study 178-CL-046	29
Table 27: Change from Baseline to Final visit in Micturitions per 24 hours by Age group – Study 178-CL-047	29
Table 28: Change from Baseline to Final visit in Micturitions per 24 hours by Age group – Study 178-CL-074	29
Table 29: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Race – Study 178-CL-047	29
Table 30: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Race – Study 178-CL-074	29
Table 31: Change from Baseline to Final visit in Micturitions per 24 hours by Race – Study 178-CL-047	30
Table 32: Change from Baseline to Final visit in Micturitions per 24 hours by Race – Study 178-CL-074	30
Table 33: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Region – Study 178-CL-074	30
Table 34: Change from Baseline to Final visit in Micturitions per 24 hours by Region – Study 178-CL-074	30
Table 35: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Beta-blockers use at baseline – Study 178-CL-046	31
Table 36: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Beta-blockers use at baseline – Study 178-CL-047	31
Table 37: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Beta-blockers use at baseline – Study 178-CL-074	31
Table 38: Change from Baseline to Final visit in micturitions per 24 hours by Beta-blockers use at baseline – Study 178-CL-046	31
Table 39: Change from Baseline to Final visit in micturitions per 24 hours by Beta-blockers use at baseline – Study 178-CL-047	32
Table 40: Change from Baseline to Final visit in micturitions per 24 hours by Beta-blockers use at baseline – Study 178-CL-074	32

LIST OF FIGURES

Figure 1: LS Mean Estimate of Change from Baseline at Each Visit in Mean Number of Incontinence Episodes per 24 Hours – Study 178-CL-074, FAS-I	17
Figure 2: LS Mean Estimate of Change from Baseline at Each Visit in Mean Number of Micturitions per 24 Hours – Study 178-CL-074, FAS.....	18
Figure 3: LS Mean Estimate of Change from Baseline in Mean Volume Voided per Micturition at Each Visit – Study 178-CL-074, FAS.....	18
Figure 4: First 3 stages in the hierarchical testing procedure.....	22

1 EXECUTIVE SUMMARY

In this submission, the Applicant is seeking approval of mirabegron 50 mg daily for treatment of overactive bladder. To support this claim, the safety and efficacy data from three Phase 3 clinical trials (178-CL-046, 178-CL-047 and 178-CL-074) were submitted. This review evaluates to determine from a statistical perspective if the submitted information supports this claim.

All three studies were multinational, multi-center, double-blind, randomized, parallel group, placebo controlled Phase 3 studies with a 12-week treatment period. Study 178-CL-046 was also an active controlled study with tolterodine SR 4 mg. Study 178-CL-046 was conducted in Europe and subjects were randomized to placebo, mirabegron 50 mg, 100 mg and tolterodine SR 4 mg; Study 178-CL-047 was conducted in US and Canada and subjects were randomized to placebo, Mirabegron 50 mg and 100 mg. Study 178-CL-074 was conducted in Europe, US and Canada and subjects were randomized to placebo, mirabegron 25 mg and 50 mg. The 25mg dose was intended for patients with severe renal or moderate hepatic impairment.

Prior to each clinical visit, subjects were instructed to complete a 3-day micturition diary to record times of micturition, void volume, urgency severity, incontinence episodes, sleep interruption and pad use.

The co-primary efficacy endpoints, based on the subject 3-day micturition diary, were:

- change from baseline to end of treatment (final visit) in mean number of micturitions per 24 hours;
- change from baseline to end of treatment (final visit) in mean number of incontinence episodes per 24 hours.

In all three studies, both co-primary efficacy endpoints were analyzed using an ANCOVA model which included treatment group, gender, geographical region, and baseline measurement. Point estimates and two-sided 95% confidence intervals for the mean change from baseline within each treatment group as well as for the difference in mean change from baseline between each mirabegron treatment group and placebo (and between tolterodine and placebo) were calculated. For the change from baseline to end of treatment (final visit) in mean number of micturitions per 24 hours, the pair-wise p-values were derived from the above ANCOVA model for the comparisons between each active treatment group vs. placebo group. Due to the non-normal data distribution of change from baseline in mean number of incontinence episodes, a stratified rank ANCOVA model was utilized for hypothesis testing and calculating the pair-wise p-values.

No statistical issues were identified in this submission. The Applicant adhered to the statistical methods for the primary and key secondary endpoints as specified in the protocol and statistical analysis plan. The data from the three Phase 3 studies demonstrated that all three mirabegron doses (25 mg, 50 mg and 100 mg) had statistically significant improvement in the pre-specified co-primary efficacy endpoints compared with placebo.

From a statistical perspective, all doses of mirabegron (25 mg, 50 mg and 100 mg) are effective in treating overactive bladder. Although mirabegron 50 mg is the proposed dose for general OAB patients by the Applicant, mirabegron 25 mg dose also showed very similar efficacy on the co-primary endpoints compared to mirabegron 50 mg dose in one Phase 3 trial with adequate sample size. Therefore, mirabegron 25 mg dose should be considered for general OAB patients as well.

2 INTRODUCTION

2.1 Overview

The Applicant, Astellas Pharma Global Development INC., seeks approval of mirabegron 50 mg once daily (and 25 mg daily for patients with severe renal or moderate hepatic impairment) for the treatment of overactive bladder (OAB).

According to the Applicant, mirabegron is a selective agonist for human beta 3-adrenoceptor (beta 3-AR). And it is a new chemical entity; first-in-class compound, which is indicated for the symptomatic treatment of urgency, increased micturitions frequency and /or urgency incontinence as may occur in patients with OAB.

The statistical review for this NDA is based on the three double-blind phase 3 studies, 178-CL-046, 178-CL-047 and 178-CL-048, which are briefly summarized in Table 1. During the development of mirabegron, the protocols for the primary phase 3 studies 178-CL-046 and 178-CL-047 were submitted to FDA for Special Protocol Assessment (SPA) in December 2007. On February 5, 2008, a regulatory letter with comments from the Division was conveyed to the Applicant. No agreement was reached by the Division. The Applicant initiated the two Phase 3 trials and submitted amended protocols and requested another SPA on March 5, 2009. In a regulatory letter dated March 16, 2009, the Division informed Astellas that an SPA cannot be provided once a study has begun. FDA sent the Applicant the comments for the amended protocols on May 20, 2009.

Table 1: List of all Studies included in the Statistical Review

Study	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
178-CL-046	Phase 3, randomized, double blind, parallel group, placebo and active-controlled	12 weeks	30 days	Randomized: Placebo: 497 Mirabegron 50 mg: 497 Mirabegron 100 mg: 498 Tolterodine SR 4mg: 495	<ul style="list-style-type: none">• ≥18 years of age• Had symptoms of overactive bladder for at least 3 months• ≥8 micturitions per 24 hours• At least 3 episodes of urgency (grade 3 or 4) with or without incontinence in a 3-day diary period• Treatment naïve or prior antimuscarinic therapy
178-CL-047	Phase 3 randomized, double blind, parallel group, placebo controlled	12 weeks	30 days	Randomized: Placebo: 454 Mirabegron 50 mg: 442 Mirabegron 100 mg: 433	
178-CL-074	Phase 3, randomized, double blind, parallel group, placebo controlled	12 weeks	2 weeks	Randomized: Placebo: 433 Mirabegron 25 mg: 433 Mirabegron 50 mg: 440	

Source: Statistical reviewer's summary.

2.2 Data Sources

The study reports, data and additional information were submitted electronically. These items are located in the Electronic Document Room at [\\Cdsub1\evsprod\NDA202611](#) under submission dates 08/30/2011, 10/11/2011 and 12/20/2011.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The Applicant submitted both the tabulation data and analysis data for the three studies. In the original submission made on August 30, 2011, the size of some data sets (analysis data and tabulation data) was extremely large to access by the FDA review team. Information request to resize the data sets was sent to the Applicant on November 3, 2011. General guidance, response to the Applicant's questions and sample SAS codes to perform the datasets resizing were sent to the Applicant by FDA as well. The Applicant resubmitted the resized data on December 15, 2011, which can be accessed by the review team. Data sets were complete and documented.

The statistical analyses of efficacy endpoints in each study were carried out following the pre-specified statistical analysis plan.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

3.2.1.1 Study Design

All three studies 178-CL-046, 178-CL-047 and 178-CL-074 were multinational, multi-center, double-blind, randomized, parallel group, placebo controlled phase III studies. Study 178-CL-046 was also an active controlled study with tolterodine SR 4 mg and it was conducted in Europe; study 178-CL-047 was conducted in US and Canada; and study 178-CL-074 was conducted in Europe, US and Canada.

Each study contained 6 visits: a screening visit, a randomization visit, three visits during the treatment period and a follow-up visit. At end of the screening visit, study subjects received medication for a single-blind 2-week placebo run-in period. Following the placebo run-in period, at Visit 2 (randomization visit), eligible subjects were randomized into a double-blind, placebo and active controlled, 12-week treatment period if they experienced frequency of micturition on an average of ≥ 8 times per 24-hour period during the 3-day micturition diary period, experienced at least 3 episodes of urgency (grade 3 or 4) with or without incontinence during the 3-day micturition diary period and met other inclusion and exclusion criteria.

In study 178-CL-046, eligible subjects were randomized to mirabegron 50 mg, mirabegron 100 mg, placebo, and tolterodine SR 4 mg at 1:1:1:1 ratio in the treatment period. In study 178-CL-047, subjects were randomized to mirabegron 50 mg, mirabegron 100 mg, and placebo at 1:1:1 ratio. In study 178-CL-074, subjects were randomized to mirabegron 25 mg, mirabegron 50 mg and placebo at 1:1:1 ratio.

3.2.1.2 Primary/Key Secondary Efficacy Endpoints

Subjects were instructed to complete the 3-day micturitions diary before each visit in the run-in and treatment periods. Times of micturition, voided volume (2 of 3 days), urgency severity, incontinence episodes and pad use were recorded in the micturition diary. The diaries were reviewed during each visit by the investigator or research nurse to ensure accuracy of completion.

The primary efficacy endpoints, based on the 3-day micturitions diary were defined as follows,

- change from baseline to end of treatment (final visit) in mean number of micturitions per 24 hours;
- change from baseline to end of treatment (final visit) in mean number of incontinence episodes per 24 hours.

The Applicant's also defined key secondary efficacy endpoints based on the 3-day micturitions diary as:

- change from baseline to end of treatment (final visit) in mean volume voided per micturitions;
- change from baseline to Week 4 in mean number of micturitions per 24 hours;
- change from baseline to Week 4 in mean number of incontinence episodes per 24 hours.

In study 178-CL-074, in addition to the above key secondary efficacy endpoints, the Applicant also defined additional key secondary efficacy endpoints as:

- change from baseline to end of treatment (final visit) in mean level of urgency;
- change from baseline to end of treatment (final visit) in mean number of urgency incontinence episodes per 24 hours;
- change from baseline to end of treatment (final visit) in mean number of urgency episodes (grades 3 or 4) per 24 hours.

For subjects who do not have a value at Week 12 for an efficacy or safety variable, LOCF methodology was utilized for deriving the final visit value.

3.2.1.3 Multiplicity Control Approach

In each of the three studies, the Applicant adopted a stepwise parallel gate-keeping procedure to control the type I error rate over multiple active treatment groups and multiple efficacy endpoints at the 0.05 significance level. In study 178-CL-046, no adjustment for multiplicity between tolterodine and placebo was done. The stepwise parallel gate-keeping procedure was performed in the following ordered stages. The first 5 stages were common in all three studies and stages 6-8 were specific for study 178-CL-074.

Stage 1: incontinence episodes at final visit

Stage 2: micturitions at final visit

Stage 3: volume voided per micturition at final visit

Stage 4: incontinence episodes at Week 4

Stage 5: micturitions at Week 4

Stage 6: mean level of urgency at final visit

Stage 7: urgency incontinence episodes at final visit

Stage 8: urgency episodes (grades 3 or 4) at final visit

In the gate-keeping procedure, only if a mirabegron dose achieved statistically significance at all previous stages, this dose group can proceed to the next stage.

Within each stage, the Hochberg procedure was used to control the overall stage Type I error rate at $\alpha = 0.05$ level for multiple treatment group comparisons. The hypothesis for a given efficacy endpoint was tested utilizing the pre-specified statistical methodology. The p-values generated from the statistical test were ordered from the largest value (p1) to the smallest value (p2).

- If p1 and p2 were > 0.05, then all mirabegron treatment group comparisons to placebo were not statistically significant.
- If the largest p-value p1 was ≤ 0.05, then all mirabegron treatment group comparisons to placebo were statistically significant.
- If the largest p-value p1 was > 0.05, then the corresponding mirabegron treatment 1 comparison to placebo was considered not statistically significant and p2 would be compared to $\alpha/2 = 0.025$.
 - If p2 was > 0.025, then the corresponding mirabegron treatment 2 comparison to placebo was considered not statistically significant,
 - If p2 was ≤ 0.025, then the corresponding mirabegron treatment 2 comparison to placebo was considered statistically significant.

Only the mirabegron treatment groups which succeeded in stage 1 would be evaluated in stage 2 for micturitions. If both mirabegron treatment groups succeeded in Stage 1 then both mirabegron treatment groups would proceed to stage 2 and the Hochberg procedure was implemented at $\alpha = 0.05$. If one mirabegron treatment group succeeds in Stage 1, then only that mirabegron treatment group proceeded to Stage 2 and the comparison versus placebo was assessed at $\alpha = 0.025$. If both mirabegron treatment groups failed in Stage 1, then neither mirabegron treatment group would proceed to subsequent Stages. The above process was repeated in the left stages. A graphical display of the hierarchical testing procedure is shown in Appendix (Figure 1) for the first 3 stages.

3.2.2 Patient Disposition, Demographic and Baseline Characteristics

3.2.2.1 Patient Disposition

The disposition of study patients are summarized by treatment groups in Table 2 to Table 4. In study 178-CL-046, a total of 1987 patients were randomized to four treatment groups and the study discontinuation rate is 9.9%, ranging from 8.9% to 11.5% across the treatment groups. In study 178-CL-047, a total of 1329 patients were randomized to three treatments and the study discontinuation rate is 13.6%, ranging from 12.2% to 15.2% across the treatment groups. In study 178-CL-047, a total of 1306 patients were randomized to three treatment groups and the study discontinuation rate is 12.7%, ranging from 10.6% to 15.2% across the treatment groups. For all studies, the most common reasons for discontinuation from the study were adverse events and withdrawal of consent.

Table 2: Summary of Subject Disposition – Study 178-CL-046

	Placebo n (%)	Mirabegron 50 mg n (%)	Mirabegron 100 mg n (%)	Tolterodine SR 4 mg n (%)
Randomized	497 (100.0)	497 (100.0)	498(100.0)	495 (100.0)
Discontinued from study	44 (8.9)	57 (11.5)	45 (9.0)	50(10.1)
Eligibility criterion not met	5 (1.0)	8(1.6)	0	4(0.8)
Adverse event	13(2.6)	25(5.0)	16(3.2)	24(4.8)
Lack of Efficacy	5(1.0)	6(1.2)	2(0.4)	3(0.6)
Withdrew consent	11(2.2)	9(1.8)	17(3.4)	9(1.8)
Lost to follow-up	4(0.8)	3(0.6)	2(0.4)	5(1.0)
Protocol violation	2(0.4)	3(0.6)	5(1.0)	3(0.6)
Randomized but never received study drug	2(0.4)	1(0.2)	1(0.2)	0
Other	2(0.4)	2(0.4)	2(0.4)	2(0.4)

Source: Figure 2 in the Applicant's study report for 178-CL-046.

Table 3: Summary of Subject Disposition – Study 178-CL-047

	Placebo n (%)	Mirabegron 50 mg n (%)	Mirabegron 100 mg n (%)
Randomized	454 (100.0)	442 (100.0)	433(100.0)
Discontinued from study	69 (15.2)	59 (13.3)	53 (12.2)
Eligibility criterion not met	0	0	1(0.2)
Adverse event	17(3.7)	18(4.1)	19(4.4)
Lack of Efficacy	9(2.0)	1(0.2)	5(1.2)
Withdrew consent	29(6.4)	22(5.0)	16(3.7)
Lost to follow-up	2(0.4)	9(2.0)	3(0.7)
Protocol violation	7(1.5)	4(0.9)	5(1.2)
Randomized but never received study drug	1(0.2)	0	0
Other	4(0.9)	5(1.1)	4(0.9)

Source: Figure 2 in the Applicant’s study report for 178-CL-047.

Table 4: Summary of Subject Disposition – Study 178-CL-074

	Placebo n (%)	Mirabegron 25 mg n (%)	Mirabegron 50mg n (%)
Randomized	433 (100.0)	433 (100.0)	440(100.0)
Discontinued from study	66 (15.2)	46 (10.6)	54 (12.3)
Eligibility criterion not met	1(0.2)	1(0.2)	0
Adverse event	15(3.5)	17(3.9)	12(2.7)
Lack of Efficacy	11(2.5)	4(0.9)	3(0.7)
Withdrew consent	20(4.6)	12(2.8)	18(4.1)
Lost to follow-up	4(0.9)	3(0.7)	3(0.7)
Protocol violation	5(1.2)	3(0.7)	8(1.8)
Randomized but never received study drug	0	1(0.2)	0
Other	10(2.3)	5(1.2)	10(2.3)

Source: Figure 2 in the Applicant’s study report for 178-CL-074.

For primary efficacy evaluation, the Applicant pre-defined the following analyses datasets in each study:

- Full analysis set (FAS): all randomized subjects who took at least one dose of double-blind study medication and had a micturition measurement in the baseline diary and at least one post-baseline visit diary with a micturition measurement.
- FAS incontinence (FAS-I): all randomized subjects who took at least one dose of double-blind study medication and had a micturition measurement and at least one incontinence episode in the baseline diary and at least one post-baseline diary with a micturition measurement.

The numbers of patients in the defined efficacy analysis sets are presented in Table 5–7.

Table 5: Summary of Efficacy analysis sets – Study 178-CL-046

Analysis Population	Placebo n (%)	Mirabegron 50 mg n (%)	Mirabegron 100 mg n (%)	Tolterodine SR 4 mg n (%)	Total n (%)
Randomized analysis set	497(100.0%)	497(100.0%)	498(100.0%)	495(100.0%)	1987(100.0%)
Full Analysis Set(FAS)	480 (96.6%)	473 (95.2%)	478 (96.0%)	475 (96.0%)	1906 (95.9%)
FAS incontinence	291 (58.6%)	293 (59.0%)	281 (56.4%)	300 (60.6%)	1165 (58.6%)

Source: Table 2 in the Applicant’s study report for 178-CL-046.

Table 6: Summary of Efficacy analysis sets – Study 178-CL-047

Analysis Population	Placebo n (%)	Mirabegron 50 mg n (%)	Mirabegron 100 mg n (%)	Total n (%)
Randomized analysis set	454 (100.0%)	442 (100.0%)	433 (100.0%)	1329 (100.0%)
Full Analysis Set(FAS)	433 (95.4%)	425 (96.2%)	412 (95.2%)	1270 (95.6%)
FAS incontinence	325 (71.6%)	312 (70.6%)	296 (68.4%)	933 (70.2%)

Source: Table 2 in the Applicant’s study report for 178-CL-047.

Table 7: Summary of Efficacy analysis sets – Study 178-CL-074

Analysis Population	Placebo n (%)	Mirabegron 25 mg n (%)	Mirabegron 50 mg n (%)	Total n (%)
Randomized analysis set	433 (100.0%)	433 (100.0%)	440 (100.0%)	1306 (100.0%)
Full Analysis Set(FAS)	415 (95.8%)	410 (94.7%)	426 (96.8%)	1251 (95.8%)
FAS incontinence	262 (60.5%)	254 (58.7%)	257 (58.4%)	773 (59.2%)

Source: Table 3 in the Applicant’s study report for 178-CL-074.

3.2.2.2 Demographics and Baseline Characteristics

The demographics and baseline characteristics of the treatment groups are summarized in the Appendix (Table 13-15) for each of the three studies. Study 178-CL-046 was conducted in Europe. Study 178-CL-047 had 90% US subjects and 10% Canadian subjects. Study 178-CL-074 had 46.4% European subjects, 46.6% US subjects and 7% Canadian subjects.

In all studies, the majority of subjects were white (99.2% for 178-CL-046; 88.2% for 178-CL-047; 90.6% for 178-CL-047), and the percentages of white subjects were comparable across treatment groups in each study. The percentages of female subjects were 72.0% for study 178-CL-046, 74.8% for 178-CL-047 and 68.5% for 178-CL-074. The mean age of subjects was 59.1 years in study 178-CL-046 and 60.2 years in 178-CL-047 and 59.1 years in 178-CL-074. Overall, the demographics across treatment groups were similar in each study.

3.2.3 Statistical Methodologies

3.2.3.1 Applicant’s analyses

Table 8 highlights the analysis datasets and key statistical methods used to evaluate the efficacy endpoints in each study. The incontinence episodes endpoint was analyzed using FAS-I dataset, while all other efficacy endpoints were analyzed using FAS dataset.

Change from baseline to a specific visit in an efficacy variable was analyzed using an ANCOVA model which included treatment group, gender, geographical region, and baseline measurement. Point estimates and two-sided 95% confidence intervals for the mean change from baseline within each treatment group as well as for the difference in mean change from baseline between each mirabegron treatment group and placebo (and between tolterodine and placebo) were calculated. If the normality assumption for the data of the efficacy endpoint was not violated, the pair-wise P-values were derived from the above ANCOVA model for the comparisons between each active treatment group vs. placebo group.

Due to the non-normally distributed data of change from baseline in mean number of incontinence episodes, a stratified rank ANCOVA was utilized for hypothesis testing and calculating the pair-wise p-values. The following steps were performed for the stratified rank ANCOVA for each pair-wise treatment group difference of interest:

1. Standardized ranks within each geographical region were derived across the 2 treatment groups for the baseline value and the change from baseline value. Standardized ranks were used to adjust for differences in the number of subjects at each geographical region.
2. With the standardized ranks of change from baseline as the response variable, an ANCOVA model was fitted separately for each geographical region including the baseline standardized ranks and gender as covariates.
3. The stratified mean score test was performed to compare the two treatment groups using the values of the residuals obtained from step 2 as scores and geographical region as a stratum. The p-value obtained was reported.

If no Week 12 diary data measurements were available (often because the subject were prematurely discontinued), the last available earlier post-baseline average of the diary data measurements within a designated visit window and post-dosing window was used as the final visit measurement.

In the Applicant’s analyses, the geographical region was defined for each study as follows:

- Study 178-CL-046: Eastern Europe, Western Europe;
- Study 178-CL-047: Northeast US, Midwest US, South US, East US and Central US, and Canada;
- Study 178-CL-074: Eastern Europe, Western Europe, Canada, Northeastern US, Midwestern US, Southern US and Western US.

Table 8: Analysis Approach for the Primary/Key Secondary Efficacy endpoints

Study (178-CL-)	Efficacy endpoint	Analysis Set	LOCF or Observed	Analysis Type
046, 047,074	Change from baseline to final visit in Mean Number of Micturitions per 24 hours	FAS	LOCF	ANCOVA
046, 047,074	Change from baseline to final visit in Mean Number of Incontinence Episodes per 24 hours	FAS-I	LOCF	Stratified Rank ANCOVA
046, 047,074	Change from baseline to final visit in Mean Volume Voided per Micturition	FAS	LOCF	ANCOVA
046, 047,074	Change from baseline to Week 4 in Mean Number of Micturitions per 24 hours	FAS	Observed	ANCOVA
046, 047,074	Change from baseline to Week 4 in Mean Number of Incontinence Episodes per 24 hours	FAS-I	Observed	Stratified Rank ANCOVA
074	Change from baseline to final visit in Mean Level of Urgency	FAS	LOCF	ANCOVA
074	Change from baseline to final visit in Mean Number of Urgency Incontinence Episodes per 24 hours	FAS-I	LOCF	Stratified rank ANCOVA
074	Change from baseline to final visit in Mean Number of Urgency (grades 3 or 4) Episodes per 24 hours	FAS	LOCF	ANCOVA

Source: Table 7, studies: 178-CL-046 SAP, 178-CL-047 SAP, and 178-CL-074 SAP.

3.2.3.2 Reviewer’s analyses

Mirabegron was intended to be marketed in US. Therefore, the treatment effect in US patients was of interest. By categorizing Europe and US into smaller regions as the Applicant did above, the analyses of efficacy endpoints may not be able to evaluate the treatment effect among all US patients. Because of the smaller number of patients by geographical location in Canada, the reviewer re-defined the region variable as Europe vs. US/Canada. For each study, the reviewer re-analyzed the co-primary and key secondary efficacy endpoints using the Applicant’s ANOCVA or stratified ANCOVA models with the re-defined region variable.

3.2.4 Results and Conclusions

Study 178-CL-046

The Applicant's analysis results for the efficacy endpoints are shown in Table 9 for study 178-CL-046. Both mirabegron 50 mg and 100 mg dose groups demonstrated statistically significant improvements in the co-primary efficacy endpoints and key secondary efficacy endpoints compared to the placebo group under the pre-specified multiplicity controlling procedure. At the final visit, compared with placebo group, the mirabegron 50 mg reduced 0.41 more episodes and mirabegron 100 mg reduced 0.29 more episodes reduction in the change from baseline in mean number of incontinence episodes per 24 hours. Compared with placebo group, the mirabegron 50 mg group reduced 0.60 more micturitions and mirabegron 100 mg reduced 0.44 more micturitions in the change from baseline in mean number of micturitions per 24 hours. The tolterodine SR 4 mg group did not demonstrate statistically significant improvements in the co-primary efficacy endpoints compared with the placebo group. No further testing on the key secondary efficacy endpoints was proceeded.

Table 9: Summary of Primary and Key Secondary Efficacy endpoints – Study 178-CL-046

	Placebo	Mirabegron 50 mg	Mirabegron 100 mg	Tolterodine SR 4 mg
Mean Number of Incontinence Episodes per 24 Hours (FAS-I) at final visit				
N	291	293	281	300
Mean	1.54	1.22	1.37	1.42
Change from baseline*	-1.17	-1.57	-1.46	-1.27
Difference vs. placebo (p-value†)		-0.41 (0.003#)	-0.29 (0.010#)	-0.10 (0.115)
Mean Number of Micturitions per 24 Hours (FAS) at final visit				
N	480	473	478	475
Mean	10.35	9.70	9.76	9.97
Change from baseline*	-1.34	-1.93	-1.77	-1.59
Difference vs. placebo (p-value‡)		-0.60 (<0.001#)	-0.44 (0.005#)	-0.25 (0.112)
Mean Volume Voided (mL) per Micturition (FAS) at final visit				
N	480	472	478	475
Mean	169.1	185.2	183.8	183.6
Change from baseline*	12.3	24.2	25.6	25.0
Difference vs. placebo (p-value‡)		11.9 (<0.001#)	13.2 (<0.001#)	12.6 (<0.001)
Mean Number of Incontinence Episodes per 24 Hours (FAS-I) at Week 4				
N	291	293	281	299
Mean	2.05	1.76	1.81	1.68
Change from baseline*	-0.65	-1.04	-1.03	-1.00
Difference vs. placebo (p-value†)		-0.39 (0.002#)	-0.38 (0.002#)	-0.35 (0.019)
Mean Number of Micturitions per 24 Hours (FAS) at Week 4				
N	479	471	477	474
Mean	10.93	10.47	10.24	10.46
Change from baseline*	-0.77	-1.16	-1.29	-1.10
Difference vs. placebo (p-value‡)		-0.40 (0.004#)	-0.52 (<0.001#)	-0.33 (0.016)

*Change from baseline was obtained from an ANCOVA model.

†Nominal P-values were from pair-wise comparisons vs. placebo within the stratified rank analysis of covariance (ANCOVA).

‡Nominal P-values were from pair-wise comparisons vs. placebo within the ANCOVA model.

#Statistically significantly superior to placebo with multiplicity adjustments at the 0.05 level.

Source: Tables 12.3.1.1, 12.3.1.2, 12.3.4.1, 12.3.4.2, 12.3.4.3 in the Applicant's study 178-CL-046 report and statistical reviewer's analyses.

Study 178-CL-047

The Applicant's analysis results for the efficacy endpoints are shown in Table 10 for study 178-CL-047. Both mirabegron 50 mg and 100 mg groups demonstrated statistically significant improvements in the co-primary efficacy endpoints and key secondary efficacy endpoints compared to the placebo group under the pre-specified multiplicity controlling procedure. At the final visit, compared with placebo group, the mirabegron 50 mg reduced 0.34 more episodes and mirabegron 100 mg reduced 0.50 more episodes in the change from baseline in mean number of incontinence episodes per 24 hours. Compared with placebo group, the mirabegron 50 mg reduced 0.61 more micturition and mirabegron 100 mg reduced 0.70 more micturition in the change from baseline in mean number of micturitions per 24 hours.

Table 10: Summary of Primary and Key Secondary Efficacy endpoints – Study 178-CL-047

Efficacy Endpoint	Placebo	Mirabegron 50 mg	Mirabegron 100 mg
Mean Number of Incontinence Episodes per 24 Hours (FAS-I) at final visit			
N	325	312	296
Mean	1.81	1.33	1.14
Change from baseline*	-1.13	-1.47	-1.63
Difference vs. placebo (p-value†)		-0.34(0.026#)	-0.50(<0.001#)
Mean Number of Micturitions per 24 Hours (FAS) at final visit			
N	433	425	412
Mean	10.51	10.09	9.91
Change from baseline*	-1.05	-1.66	-1.75
Difference vs. placebo (p-value‡)		-0.61(0.001#)	-0.70(< 0.001#)
Mean Volume Voided (mL) per Micturition (FAS) at final visit			
N	433	424	412
Mean	164.6	174.4	175.4
Change from baseline*	7.0	18.2	18.0
Difference vs. placebo (p-value‡)		11.1(0.001#)	11.0(0.002#)
Mean Number of Incontinence Episodes per 24 Hours (FAS-I) at Week 4			
N	325	309	293
Mean	2.21	1.59	1.58
Change from baseline*	-0.72	-1.20	-1.18
Difference vs. placebo (p-value†)		-0.48(0.003#)	-0.46(<0.001#)
Mean Number of Micturitions per 24 Hours (FAS) at Week 4			
N	433	422	409
Mean	10.79	10.58	10.29
Change from baseline*	-0.77	-1.19	-1.37
Difference vs. placebo (p-value‡)		-0.42(0.022#)	-0.60(0.001#)

*Change from baseline was obtained from an ANCOVA model.

†Nominal P-values were from pairwise comparisons vs. placebo within the stratified rank analysis of covariance (ANCOVA).

‡Nominal P-values were from pairwise comparisons vs. placebo within the ANCOVA model.

#Statistically significantly superior to placebo with multiplicity adjustments at the 0.05 level.

Source: Table 12.3.1.1, 12.3.1.2, 12.3.4.1, 12.3.4.2 and 12.3.4.3 in the Applicant's study 178-CL-047 report.

Study 178-CL-074

The Applicant's analysis results for the efficacy endpoints are shown in Table 11 for study 178-CL-074. Both mirabegron 25 mg and 50 mg groups demonstrated statistically significant improvements in the co-

primary efficacy endpoints compared to the placebo group under the pre-specified multiplicity controlling procedure. At the final visit, compared with placebo group, the mirabegron 25 mg group had 0.40 more episode reduction and mirabegron 50 mg group had 0.42 more episode reduction in change from baseline in mean number of incontinence episodes per 24 hours. And, compared with placebo group, the mirabegron 25 mg group had 0.47 more micturition reduction and mirabegron 50 mg group had 0.42 more micturition reduction in change from baseline in mean number of micturitions per 24 hours.

Mirabegron 25 mg group did not demonstrate statistically significance on all key secondary efficacy endpoints and the 50 mg dose group showed statistically significant improvements in the key secondary efficacy endpoints: change from baseline to final visit in mean volume voided (mL) per micturition and change from baseline to Week 4 in mean number of incontinence episodes per 24 hours. The mirabegron 50 mg group had 12.4 mL increase on the change from baseline to the final visit in mean volume voided per micturitions compared with placebo.

Table 11: Summary of Primary and Key Secondary Efficacy endpoints – Study 178-CL-074

Efficacy Endpoint	Placebo	Mirabegron 25 mg	Mirabegron 50 mg
Mean Number of Incontinence Episodes per 24 Hours (FAS-I) at final visit			
N	262	254	257
Mean	1.54	1.21	1.13
Change from baseline*	-0.96	-1.36	-1.38
Difference vs. placebo (p-value†)		-0.40(0.005#)	-0.42 (0.001#)
Mean Number of Micturitions per 24 Hours (FAS) at final visit			
N	415	410	426
Mean	10.33	10.02	10.04
Change from baseline*	-1.18	-1.65	-1.60
Difference vs. placebo (p-value‡)		-0.47(0.007#)	-0.42(0.015#)
Mean Volume Voided (mL) per Micturition (FAS) at final visit			
N	415	410	426
Mean	172.3	177.6	180.3
Change from baseline*	8.3	12.8	20.7
Difference vs. placebo (p-value‡)		4.6(0.15)	12.4(<0.001#)
Mean Number of Incontinence Episodes per 24 Hours (FAS-I) at Week 4			
N	262	254	255
Mean	1.87	1.62	1.40
Change from baseline*	-0.62	-0.96	-1.13
Difference vs. placebo (p-value†)		-0.34(0.039)	-0.51(<0.001#)
Mean Number of Micturitions per 24 Hours (FAS) at Week 4			
N	415	410	424
Mean	10.73	10.71	10.52
Change from baseline*	-0.78	-0.96	-1.14
Difference vs. placebo (p-value‡)		-0.18 (0.30)	-0.37(0.035)

*Change from baseline was obtained from an ANCOVA model.

†Nominal P-values were from pairwise comparisons vs. placebo within the stratified rank analysis of covariance (ANCOVA).

‡Nominal P-values were from pairwise comparisons vs. placebo within the ANCOVA model.

#Statistically significantly superior to placebo with multiplicity adjustments at the 0.05 level.

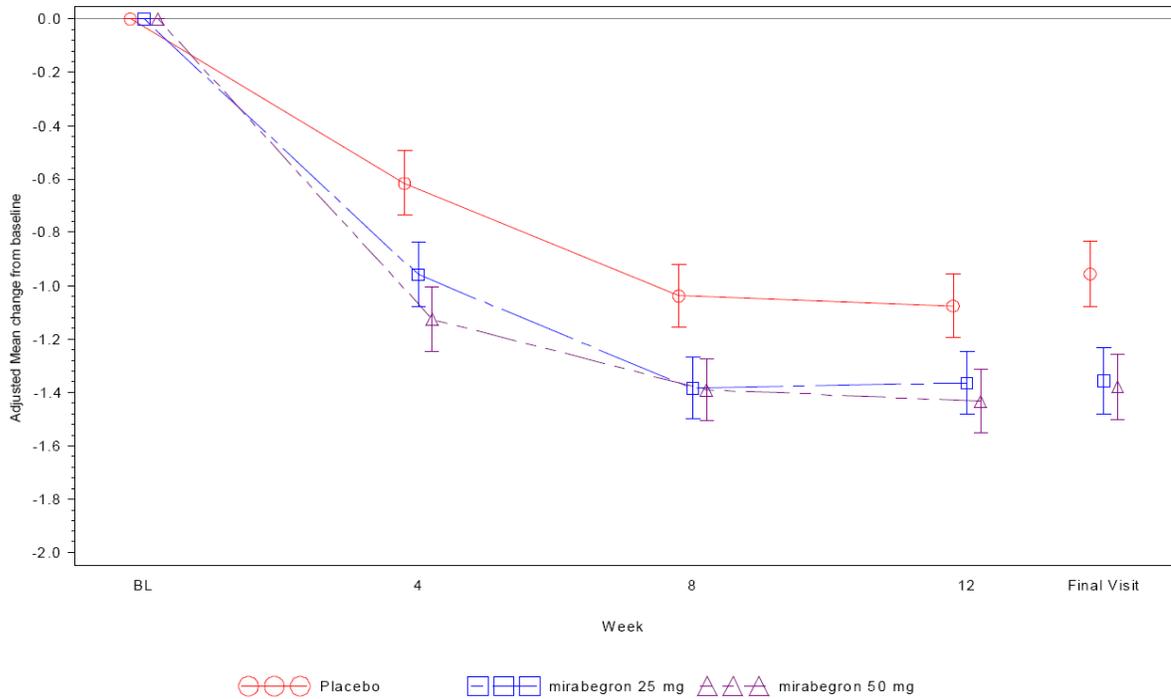
Source: Table 12.3.1.1, 12.3.1.2, 12.3.4.1, 12.3.4.2, and 12.3.4.3 in the Applicant's study 178-CL-074 report.

The analysis results of additional key secondary efficacy endpoints defined in 178-CL-074 was presented in Table 16.

The efficacy analysis results shown in Table 9-11 were validated by the reviewer independently using the Applicant's submitted data.

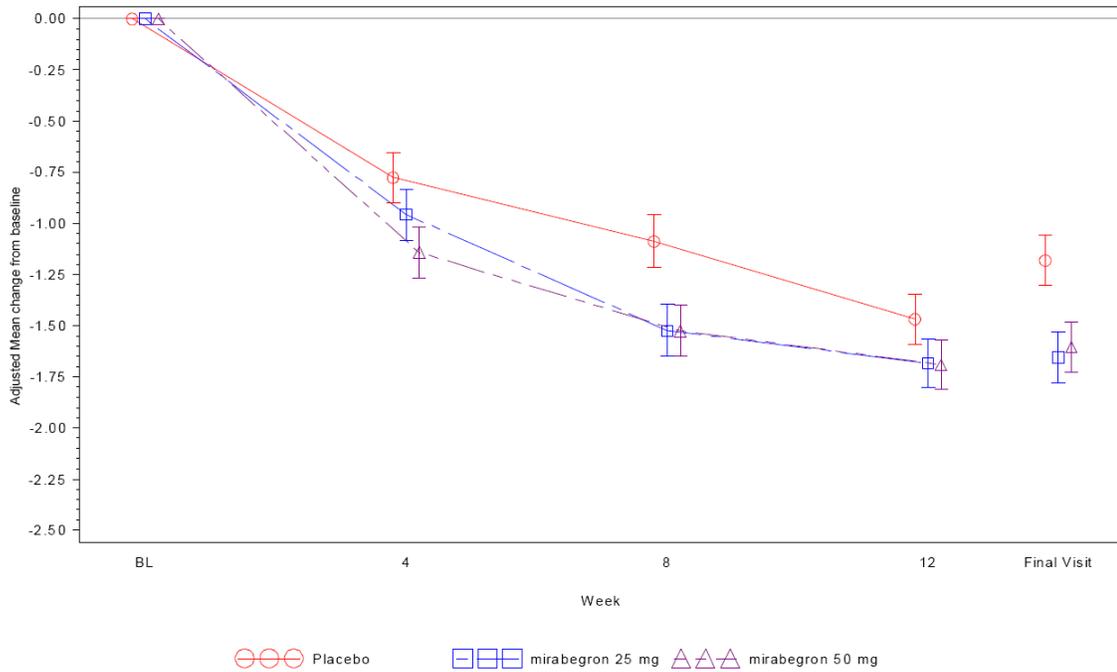
Although mirabegron 50 mg is the proposed dose for general OAB patients by the Applicant, mirabegron 25 mg also showed very similar efficacy on the co-primary endpoints compared to mirabegron 50 mg in Study 178-CL-074. Figure 1 to 3 demonstrate the Least square (LS) mean estimate of change from baseline over time at each visit in mean number of incontinence episodes per 24 hours, mean number of micturitions per 24 hours and mean volume voided per micturition.

Figure 1: LS Mean Estimate of Change from Baseline at Each Visit in Mean Number of Incontinence Episodes per 24 Hours – Study 178-CL-074, FAS-I



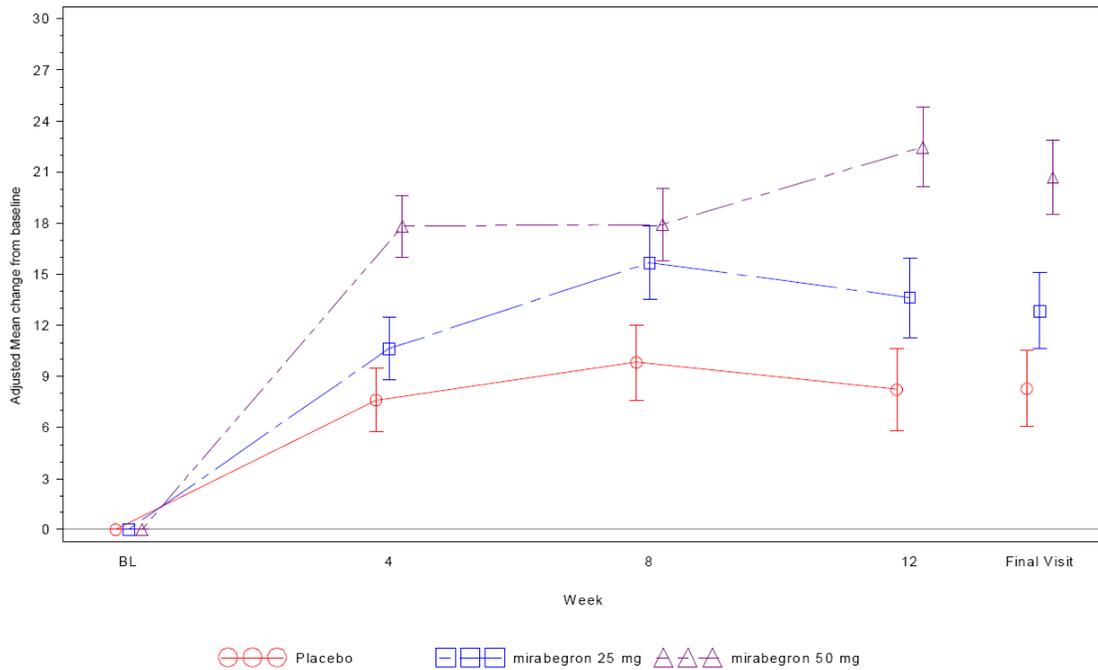
Source: Figure 3 in the study 178-CL-074 report.

Figure 2: LS Mean Estimate of Change from Baseline at Each Visit in Mean Number of Micturitions per 24 Hours – Study 178-CL-074, FAS



Source: Figure 4 in the study 178-CL-074 report.

Figure 3: LS Mean Estimate of Change from Baseline in Mean Volume Voided per Micturition at Each Visit – Study 178-CL-074, FAS



Source: Figure 5 in the study 178-CL-074 report.

The LS mean change from baseline in the incontinence episodes per 24 hours and mean number of micturitions per 24 hours were very similar at Week 8, and maintained to Week 12. For the mean volume voided per micturition, the adjusted mean change from baseline had the smallest difference between 25 mg and 50 mg dose groups at Week 8 compared to Week 4 and Week 12 or the final visit.

Reviewer’s comments:

1. All three doses of mirabegron (25, 50, and 100 mg) were statistically superior to placebo with regards to co-primary endpoints, although 25 mg dose was effective starting at Week 8 and maintained similar efficacy at Week 12. The treatment effect of 25 mg dose was comparable to the treatment effects of mirabegron 50 mg and 100 mg.
2. There was no treatment difference between US/Canadian and European patients based on FDA analysis when larger regions were categorized above as opposed to the Applicant’s analysis based on smaller regions. This evaluation was made to verify the treatment effect in the US, where this drug was intended to be marketed.

3.3 Evaluation of Safety

Refer to the clinical reviewer’s report for evaluation of safety data.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

Efficacy of mirabegron was also evaluated by subgroups defined by gender, age, race and geographical region. The categories for each subgroup variable are defined in the following table. In the subgroup analyses, the categories of region were re-defined by the reviewer.

Table 12: Subgroup categories defined in each study

Grouping variable	Subgroups
Gender	Female Male
Age group	< 65 years >= 65 years
Race	White Non-white
Region	Europe, US/Canada

The subgroup by treatment group interactions for the two co-primary efficacy endpoints were analyzed using ANCOVA models which include treatment group, gender, subgroup variable and subgroup by treatment interaction as fixed factors and baseline as a covariate.

The LS mean estimates for mean changes from baseline, SE’s, two-sided 95% CIs within each treatment group as well as the difference in adjusted mean change from baseline and corresponding two-sided 95% CIs between each treatment group and the placebo group for each subgroup level were calculated. In

addition, the corresponding p-value for subgroup by treatment interaction term was calculated from the model.

4.1.1 Gender

The treatment effect of mirabegron 50 mg dose on the change from baseline in mean number of incontinence episodes per 24 hours was similar in males and females for study 178-CL-046 and 178-CL-074. In study 178-CL-047, due to the extreme large placebo response in males, mirabegron 50 mg didn't show any treatment effect compared to placebo on the incontinence episodes reduction.

With regards to the change from baseline in mean number of micturitions per 24 hours, the treatment by subgroup interaction was not statistically significant, but the treatment effect of mirabegron 50 mg dose was numerically slightly better in females than in males.

4.1.2 Age group

In all three studies, the treatment effect on the change from baseline in incontinence episodes per 24 hours in mirabegron groups were numerically larger in subjects who were ≥ 65 years old than subjects who were < 65 years old.

Except the mirabegron 50 mg group in study 178-CL-046, in all three studies the treatment effect on the change from baseline in micturitions per 24 hours in mirabegron groups were numerically larger in the subjects who were ≥ 65 years old than in subjects who were < 65 years old.

4.1.3 Race

More than 99% of subjects in study 178-CL-046 were white; therefore no subgroup analysis by race was done. In studies 178-CL-047 and 178-CL-074, about 10% of subjects were non-white. Mirabegron had treatment effect in white subjects on change from baseline in mean number of incontinence episodes but not in the non-white subjects. Similar pattern was also seen on the treatment effect in micturition reduction. This may due to the large placebo effect observed in non-white subjects compared to the white subjects in the two studies.

4.1.4 Region

As study 178-CL-046 was conducted in Europe and study 178-CL-047 was conducted in US/Canada, the reviewer's subgroup analysis of co-primary efficacy endpoints by region was only conducted for study 178-CL-074, for which the region was categorized as Europe and US/Canada. The treatment effect of mirabegron on the co-primary efficacy endpoints was consistent across regions in study 178-CL-074.

4.2 Other Special/Subgroup Populations

The impact of beta-blockers use on the treatment effect of mirabegron was of clinical interest. To explore this, the statistical reviewer conducted subgroup analysis of the co-primary endpoints by beta-blockers use status at baseline (Yes, No) using the same ANCOVA model as other subgroup analyses. No consistent impact of beta-blockers use was observed on the treatment effect of mirabegron with respect to the co-primary efficacy endpoints across studies.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The Applicant submitted three double-blind phase 3 studies (178-CL-046, 178-CL-047 and 178-CL-074) to demonstrate superiority of mirabegron compared to placebo. Study 178-CL-046 and 178-CL-047 were designed to evaluate mirabegron 50 mg and 100 mg doses, and study 178-CL-074 was designed to evaluate mirabegron 25 mg and 50 mg doses. All three studies were designed to demonstrate efficacy with respect to two co-primary endpoints.

Although mirabegron 50 mg was the proposed dose for general OAB patients by the Applicant, mirabegron 25 mg also showed very similar efficacy on the co-primary endpoints compared to mirabegron 50 mg dose in one Phase 3 trial with adequate sample size.

5.2 Conclusions and Recommendations

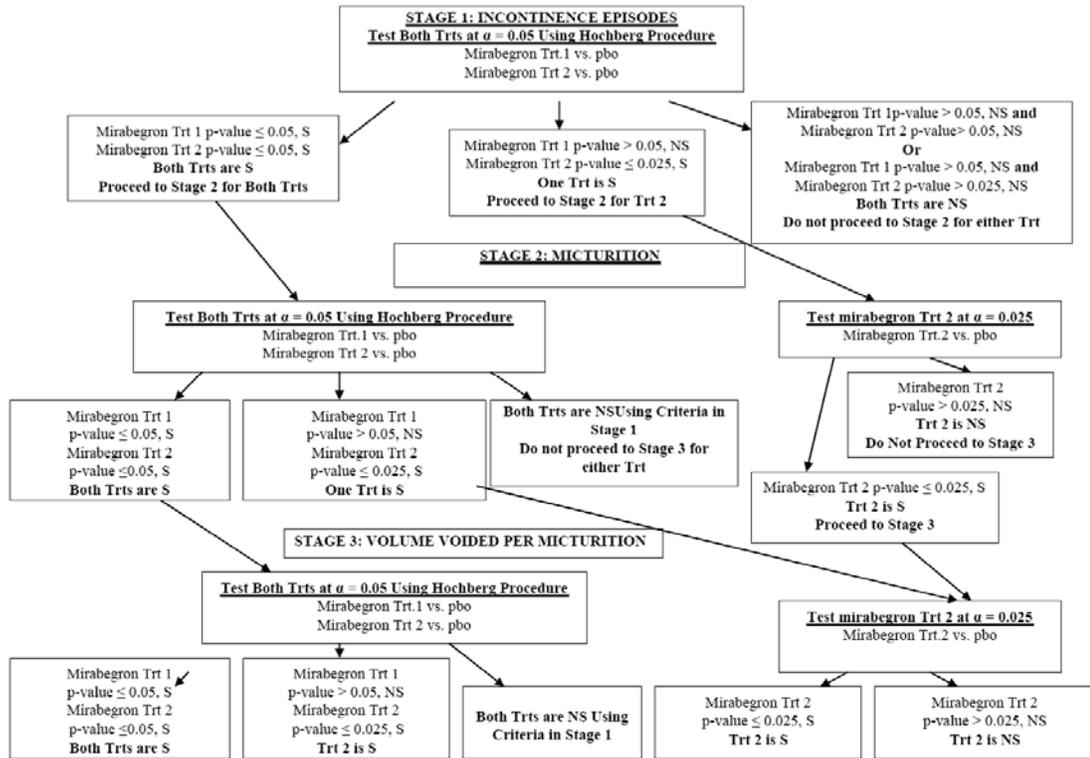
The purpose of this review is to evaluate the efficacy data in support of mirabegron in the treatment of OAB. Based on reviewer's analyses, the results supported the efficacy of mirabegron 25 mg, 50 mg and 100 mg in the improvement of all two protocol specified co-primary endpoints. The treatment effects of both mirabegron 25 mg and 50 mg dose were very similar on both co-primary endpoints.

From a statistical perspective, all doses of mirabegron (25 mg 50 mg, and 100 mg) were effective in treating OAB. Based on the efficacy analyses, mirabegron 25 mg should be considered for general OAB patients as well.

APPENDICES

Multiplicity control

Figure 4: First 3 stages in the hierarchical testing procedure.



Source: Page 43 of study 178-CL-046 SAP

Demographics and Baseline Characteristics

Table 13 Summary of Patient Demographics and Baseline Characteristics - Study 178-CL-046, SAF

Parameter	Placebo (n=494)	Mirabegron		Tolterodine SR 4 mg (n=495)	Total (n=1978)
		50 mg (n=493)	100 mg (n=496)		
Sex (n, %)					
Male	138 (27.9%)	136 (27.6%)	141 (28.4%)	134 (27.1%)	549 (27.8%)
Female	356 (72.1%)	357 (72.4%)	355 (71.6%)	361 (72.9%)	1429 (72.2%)
Age (years)					
Mean (SD)	59.2 (12.30)	59.1 (12.36)	59.0 (12.71)	59.1 (12.89)	59.1 (12.56)
Age group (years) (n, %)					
< 65	313 (63.4%)	315 (63.9%)	313 (63.1%)	303 (61.2%)	1244 (62.9%)
≥ 65	181 (36.6%)	178 (36.1%)	183 (36.9%)	192 (38.8%)	734 (37.1%)
< 75	450 (91.1%)	447 (90.7%)	450 (90.7%)	458 (92.5%)	1805 (91.3%)
≥ 75	44 (8.9%)	46 (9.3%)	46 (9.3%)	37 (7.5%)	173 (8.7%)
Race (n, %)					
White	490 (99.2%)	488 (99.0%)	492 (99.2%)	490 (99.0%)	1960 (99.1%)
Black or African American	2 (0.4%)	1 (0.2%)	1 (0.2%)	3 (0.6%)	7 (0.4%)
Asian	0	2 (0.4%)	2 (0.4%)	2 (0.4%)	6 (0.3%)
Other †	2 (0.4%)	2 (0.4%)	1 (0.2%)	0	5 (0.3%)
BMI (kg/m ²)					
n	493	493	495	495	1976
Mean (SD)	27.8 (4.96)	27.5 (4.86)	28.0 (4.95)	27.8 (4.96)	27.8 (4.93)
Geographical region (n, %)					
Eastern Europe	225 (45.5%)	222 (45.0%)	224 (45.2%)	226 (45.7%)	897 (45.3%)
Western Europe‡	269 (54.5%)	271 (55.0%)	272 (54.8%)	269 (54.3%)	1081 (54.7%)

Source: Table 5 in the Applicant's study 178-CL-046 report.

Table 14 Summary of Patient Demographics and Baseline Characteristics - Study 178-CL-047, SAF

Parameter	Placebo (n=453)	Mirabegron		Total (n=1328)
		50 mg (n=442)	100 mg (n=433)	
Sex (n, %)				
Male	108 (23.8%)	120 (27.1%)	113 (26.1%)	341 (25.7%)
Female	345 (76.2%)	322 (72.9%)	320 (73.9%)	987 (74.3%)
Age (years)				
Mean (SD)	60.1 (13.79)	59.2 (13.53)	61.0 (13.25)	60.1 (13.54)
Age group (years) (n, %)				
< 65	273 (60.3%)	274 (62.0%)	253 (58.4%)	800 (60.2%)
≥ 65	180 (39.7%)	168 (38.0%)	180 (41.6%)	528 (39.8%)
< 75	385 (85.0%)	382 (86.4%)	360 (83.1%)	1127 (84.9%)
≥ 75	68 (15.0%)	60 (13.6%)	73 (16.9%)	201 (15.1%)
Race (n, %)				
White	395 (87.2%)	391 (88.5%)	381 (88.0%)	1167 (87.9%)
Black or African American	47 (10.4%)	32 (7.2%)	37 (8.5%)	116 (8.7%)
Asian	6 (1.3%)	12 (2.7%)	8 (1.8%)	26 (2.0%)
Other	5 (1.1%) †	7 (1.6%) ‡	7 (1.6%) §	19 (1.4%)
Ethnicity (n, %)				
Hispanic/Latino	26 (5.7%)	23 (5.2%)	32 (7.4%)	81 (6.1%)
Non-Hispanic/Non-Latino	427 (94.3%)	419 (94.8%)	401 (92.6%)	1247 (93.9%)
BMI (kg/m ²)				
n	452	442	433	1327
Mean (SD)	30.4 (7.36)	30.0 (6.59)	30.2 (7.06)	30.2 (7.01)
Geographical region (n, %)				
Northeastern US	78 (17.2%)	73 (16.5%)	77 (17.8%)	228 (17.2%)
Midwestern US	61 (13.5%)	59 (13.3%)	53 (12.2%)	173 (13.0%)
Southern US	156 (34.4%)	147 (33.3%)	147 (33.9%)	450 (33.9%)
Western US	116 (25.6%)	116 (26.2%)	113 (26.1%)	345 (26.0%)
Canada	42 (9.3%)	47 (10.6%)	43 (9.9%)	132 (9.9%)

Source: Table 5 in the Applicant's study 178-CL-047 report.

Table 15 Summary of Patient Demographics and Baseline Characteristics - Study 178-CL-074, SAF

Parameter	Placebo (n = 433)	Mirabegron		Total (n = 1305)
		25 mg (n = 432)	50 mg (n = 440)	
Sex (n, %)				
Male	132 (30.5%)	139 (32.2%)	137 (31.1%)	408 (31.3%)
Female	301 (69.5%)	293 (67.8%)	303 (68.9%)	897 (68.7%)
Age (years)	58.2	58.5	60.3	59.0
Mean (SD)	(13.73)	(12.85)	(12.22)	(12.97)
Age group (years) (n, %)				
< 65	273 (63.0%)	278 (64.4%)	272 (61.8%)	823 (63.1%)
≥ 65	160 (37.0%)	154 (35.6%)	168 (38.2%)	482 (36.9%)
< 75	388 (89.6%)	400 (92.6%)	392 (89.1%)	1180 (90.4%)
≥ 75	45 (10.4%)	32 (7.4%)	48 (10.9%)	125 (9.6%)
Race (n, %)				
White	389 (89.8%)	394 (91.2%)	400 (90.9%)	1183 (90.7%)
Black or African American	35 (8.1%)	32 (7.4%)	33 (7.5%)	100 (7.7%)
Asian	7 (1.6%)	5 (1.2%)	5 (1.1%)	17 (1.3%)
Other	2 (0.5%)†	1 (0.2%)‡	2 (0.5%)§	5 (0.4%)
Ethnicity (n, %)				
Hispanic/Latino	23 (5.3%)	24 (5.6%)	21 (4.8%)	68 (5.2%)
Non-Hispanic/Non-Latino	410 (94.7%)	408 (94.4%)	419 (95.2%)	1237 (94.8%)
BMI (kg/m ²)				
n	433	432	440	1305
Mean (SD)	29.2 (6.29)	29.8 (6.50)	29.5 (6.54)	29.5 (6.45)
Geographical region (n, %)				
Eastern Europe	75 (17.3%)	76 (17.6%)	75 (17.0%)	226 (17.3%)
Western Europe	126 (29.1%)	121 (28.0%)	120 (27.3%)	367 (28.1%)
Northeastern US	42 (9.7%)	40 (9.3%)	44 (10.0%)	126 (9.7%)
Midwestern US	23 (5.3%)	25 (5.8%)	23 (5.2%)	71 (5.4%)
Southern US	70 (16.2%)	74 (17.1%)	76 (17.3%)	220 (16.9%)
Western US	66 (15.2%)	71 (16.4%)	70 (15.9%)	207 (15.9%)
Canada	31 (7.2%)	25 (5.8%)	32 (7.3%)	88 (6.7%)

Source: Table 7 in the Applicant's study 178-CL-074 report.

Applicant's efficacy analysis results

Table 16: Summary of Additional Key Secondary Efficacy endpoints - Study 178-CL-074

Efficacy Endpoint	Placebo	Mirabegron 25 mg	Mirabegron 50 mg
Mean Level of Urgency (FAS) at Final Visit			
N	413	410	426
Mean	2.21	2.15	2.11
Change from baseline*	-0.15	-0.22	-0.29
Difference vs. placebo (p-value†)		-0.07(0.083)	-0.14(<0.001)
Mean Number of Urgency Incontinence Episodes per 24 hours (FAS-I) at Final Visit			
N	256	247	251
Mean	1.36	1.08	1.00
Change from baseline*	-0.95	-1.31	-1.33
Difference vs. placebo (p-value†)		-0.36(0.004)	-0.39(0.002)
Mean Number of Episodes with Urgency (Grade 3 or Grade 4) per 24 hours (FAS) at Final Visit			
N	413	410	426
Mean	4.12	3.90	3.79
Change from baseline*	-1.35	-1.68	-1.94
Difference vs. placebo (p-value†)		-0.33(0.13)	-0.59(0.007)

†Nominal P-values were from pairwise comparisons vs. placebo within the stratified rank analysis of covariance (ANCOVA).

‡Nominal P-values were from pairwise comparisons vs. placebo within the ANCOVA model.

#Statistically significantly superior to placebo with multiplicity adjustments at the 0.05 level.

Source: Table 12.3.4.4, 12.3.4.5 and 12.3.4.6 in the Applicant's study 178-CL-074 report.

Subgroup analysis results

Table 17: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Gender – Study 178-CL-046

Gender	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Male	Placebo	48	1.45	-0.81			0.7206
	Mirabegron 50 mg	55	2.01	-1.67	-0.59	0.1195	
	Mirabegron 100 mg	42	1.71	-1.01	-0.08	0.8537	
	Tolterodine SR 4 mg	48	1.95	-1.07	-0.02	0.9536	
Female	Placebo	243	2.91	-1.19			
	Mirabegron 50 mg	238	3.02	-1.60	-0.36	0.0399	
	Mirabegron 100 mg	239	3.09	-1.60	-0.33	0.0650	
	Tolterodine SR 4 mg	252	2.76	-1.24	-0.12	0.4895	

Source: Statistical reviewer's calculation

Table 18: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Gender – Study 178-CL-047

Gender	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Male	Placebo	55	2.94	-2.02			0.0025
	Mirabegron 50 mg	61	2.46	-1.01	0.78	0.0360	
	Mirabegron 100 mg	52	2.26	-1.30	0.39	0.3123	
Female	Placebo	270	3.04	-1.05			
	Mirabegron 50 mg	251	2.85	-1.55	-0.59	0.0009	
	Mirabegron 100 mg	244	2.79	-1.61	-0.69	0.0001	

Source: Statistical reviewer's calculation

Table 19: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Gender – Study 178-CL-074

Gender	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Male	Placebo	51	1.88	-0.37			0.1135
	Mirabegron 25 mg	55	2.38	-1.70	-1.04	0.0068	
	Mirabegron 50 mg	52	2.25	-1.04	-0.45	0.2428	
Female	Placebo	211	2.56	-1.01			
	Mirabegron 25 mg	199	2.72	-1.36	-0.26	0.1834	
	Mirabegron 50 mg	205	2.57	-1.45	-0.44	0.0231	

Source: Statistical reviewer's calculation

Table 20: Change from Baseline to Final visit in Micturitions episodes per 24 hours by Gender – Study 178-CL-046

Gender	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Male	Placebo	134	11.74	-0.99			0.0517
	Mirabegron 50 mg	133	11.91	-1.51	-0.48	0.1068	
	Mirabegron 100 mg	138	11.09	-1.72	-0.91	0.0019	
	Tolterodine SR 4 mg	129	11.57	-1.55	-0.61	0.0390	
Female	Placebo	346	11.70	-1.52			
	Mirabegron 50 mg	340	11.54	-2.12	-0.64	0.0005	
	Mirabegron 100 mg	340	11.68	-1.76	-0.25	0.1758	
	Tolterodine SR 4 mg	346	11.54	-1.58	-0.11	0.5445	

Source: Statistical reviewer's calculation

Table 21: Change from Baseline to Final visit in Micturitions episodes per 24 hours by Gender – Study 178-CL-047

Gender	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Male	Placebo	101	12.08	-1.15			0.5119
	Mirabegron 50 mg	116	12.73	-1.77	-0.37	0.3286	
	Mirabegron 100 mg	103	12.05	-1.46	-0.32	0.4098	
Female	Placebo	332	11.34	-0.95			
	Mirabegron 50 mg	309	11.44	-1.69	-0.70	0.0014	
	Mirabegron 100 mg	309	11.52	-1.85	-0.82	0.0002	

Source: Statistical reviewer's calculation

Table 22: Change from Baseline to Final visit in Micturitions episodes per 24 hours by Gender – Study 178-CL-074

Gender	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Male	Placebo	127	11.30	-0.71			0.6989
	Mirabegron 25 mg	134	12.04	-1.43	-0.49	0.1168	
	Mirabegron 50 mg	133	11.56	-1.04	-0.25	0.4336	
Female	Placebo	288	11.56	-1.35			
	Mirabegron 25mg	276	11.50	-1.77	-0.44	0.0384	
	Mirabegron 50 mg	293	11.71	-1.89	-0.49	0.0192	

Source: Statistical reviewer's calculation

Table 23: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Age group – Study 178-CL-046

Age	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
<65	Placebo	177	2.60	-1.12			0.0806
	Mirabegron 50 mg	179	2.59	-1.32	-0.21	0.3156	
	Mirabegron 100 mg	173	2.90	-1.34	-0.09	0.6548	
	Tolterodine SR 4 mg	175	2.48	-1.23	-0.17	0.4093	
≥65	Placebo	114	2.78	-1.14			
	Mirabegron 50 mg	114	3.21	-2.09	-0.73	0.0047	
	Mirabegron 100 mg	108	2.87	-1.79	-0.60	0.0209	
	Tolterodine SR 4 mg	125	2.84	-1.19	-0.02	0.9465	

Source: Statistical reviewer's calculation

Table 24: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Age group – Study 178-CL-047

Age	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
<65	Placebo	191	3.10	-1.35			0.5972
	Mirabegron 50 mg	179	2.97	-1.50	-0.21	0.3132	
	Mirabegron 100 mg	169	2.57	-1.50	-0.40	0.0605	
≥65	Placebo	134	2.93	-1.03			
	Mirabegron 50 mg	133	2.51	-1.36	-0.52	0.0343	
	Mirabegron 100 mg	127	2.85	-1.64	-0.65	0.0096	

Source: Statistical reviewer's calculation

Table 25: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Age group – Study 178-CL-074

Age	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
<65	Placebo	165	2.46	-1.04			0.3687
	Mirabegron 25 mg	155	2.71	-1.48	-0.29	0.1918	
	Mirabegron 50 mg	149	2.45	-1.29	-0.25	0.2526	
≥65	Placebo	97	2.37	-0.62			
	Mirabegron 25 mg	99	2.55	-1.37	-0.65	0.0219	
	Mirabegron 50 mg	108	2.59	-1.49	-0.74	0.0075	

Source: Statistical reviewer's calculation

Table 26: Change from Baseline to Final visit in Micturitions per 24 hours by Age group – Study 178-CL-046

Age	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
<65	Placebo	302	11.82	-1.43			0.7744
	Mirabegron 50 mg	302	11.89	-2.13	-0.69	0.0005	
	Mirabegron 100 mg	306	11.63	-1.78	-0.41	0.0371	
	Tolterodine SR 4 mg	291	11.47	-1.62	-0.29	0.1498	
≥65	Placebo	178	11.53	-1.27			
	Mirabegron 50 mg	171	11.22	-1.61	-0.43	0.0926	
	Mirabegron 100 mg	172	11.29	-1.68	-0.48	0.0612	
	Tolterodine SR 4 mg	184	11.67	-1.51	-0.19	0.4474	

Source: Statistical reviewer's calculation

Table 27: Change from Baseline to Final visit in Micturitions per 24 hours by Age group – Study 178-CL-047

Age	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
<65	Placebo	261	11.60	-1.17			0.4855
	Mirabegron 50 mg	261	11.99	-1.86	-0.55	0.0220	
	Mirabegron 100 mg	244	11.68	-1.71	-0.52	0.0352	
≥65	Placebo	172	11.38	-0.74			
	Mirabegron 50 mg	164	11.49	-1.47	-0.71	0.0186	
	Mirabegron 100 mg	168	11.62	-1.80	-0.97	0.0011	

Source: Statistical reviewer's calculation

Table 28: Change from Baseline to Final visit in Micturitions per 24 hours by Age group – Study 178-CL-074

Age	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
<65	Placebo	261	11.66	-1.39			0.2935
	Mirabegron 25 mg	263	11.87	-1.68	-0.26	0.2350	
	Mirabegron 50 mg	262	11.89	-1.69	-0.25	0.2559	
≥65	Placebo	154	11.19	-0.76			
	Mirabegron 25 mg	147	11.33	-1.63	-0.80	0.0063	
	Mirabegron 50 mg	164	11.30	-1.51	-0.69	0.0150	

Source: Statistical reviewer's calculation

Table 29: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Race – Study 178-CL-047

Race	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
White	Placebo	289	3.04	-1.12			0.1697
	Mirabegron 50 mg	278	2.78	-1.38	-0.39	0.0225	
	Mirabegron 100 mg	263	2.69	-1.56	-0.61	0.0004	
Non-White	Placebo	36	2.93	-2.01			
	Mirabegron 50 mg	34	2.73	-1.92	-0.00	0.9935	
	Mirabegron 100 mg	33	2.76	-1.59	0.35	0.4648	

Source: Statistical reviewer's calculation

Table 30: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Race – Study 178-CL-074

Race	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
White	Placebo	229	2.45	-0.86			0.0397
	Mirabegron 25 mg	231	2.58	-1.44	-0.49	0.0075	
	Mirabegron 50 mg	236	2.54	-1.49	-0.57	0.0018	
Non-White	Placebo	33	2.29	-1.03			
	Mirabegron 25 mg	23	3.29	-1.43	0.18	0.7400	
	Mirabegron 50 mg	21	2.17	-0.86	0.90	0.1024	

Source: Statistical reviewer's calculation

Table 31: Change from Baseline to Final visit in Micturitions per 24 hours by Race – Study 178-CL-047

Race	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
White	Placebo	378	11.45	-0.88			0.09196
	Mirabegron 50 mg	378	11.77	-1.66	-0.67	0.0008	
	Mirabegron 100 mg	364	11.43	-1.72	-0.85	<.0001	
Non-White	Placebo	55	11.95	-1.81			
	Mirabegron 50 mg	47	11.98	-2.11	-0.28	0.6050	
	Mirabegron 100 mg	48	13.41	-1.98	0.40	0.4670	

Source: Statistical reviewer's calculation

Table 32: Change from Baseline to Final visit in Micturitions per 24 hours by Race – Study 178-CL-074

Race	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
White	Placebo	372	11.50	-1.09			0.0335
	Mirabegron 25 mg	373	11.73	-1.64	-0.49	0.0081	
	Mirabegron 50 mg	389	11.73	-1.71	-0.55	0.0025	
Non-White	Placebo	43	11.36	-1.71			
	Mirabegron 25 mg	37	11.10	-1.88	-0.25	0.6624	
	Mirabegron 50 mg	37	10.95	-0.71	0.91	0.1061	

Source: Statistical reviewer's calculation

Table 33: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Region – Study 178-CL-074

Region	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Europe	Placebo	114	2.25	-0.71			0.7322
	Mirabegron 25 mg	108	2.85	-1.45	-0.40	0.1364	
	Mirabegron 50 mg	112	2.72	-1.55	-0.56	0.0328	
US/Canada	Placebo	148	2.56	-1.02			
	Mirabegron 25mg	146	2.50	-1.42	-0.44	0.0571	
	Mirabegron 50 mg	145	2.34	-1.24	-0.35	0.1343	

Source: Statistical reviewer's calculation

Table 34: Change from Baseline to Final visit in Micturitions per 24 hours by Region – Study 178-CL-074

Region	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Europe	Placebo	196	11.68	-1.43			0.4074
	Mirabegron 25 mg	192	11.89	-1.72	-0.22	0.3777	
	Mirabegron 50 mg	193	11.81	-1.84	-0.38	0.1370	
US/Canada	Placebo	219	11.30	-0.91			
	Mirabegron 25mg	218	11.49	-1.61	-0.67	0.0057	
	Mirabegron 50 mg	233	11.55	-1.44	-0.46	0.0535	

Source: Statistical reviewer's calculation

Table 35: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Beta-blockers use at baseline – Study 178-CL-046

Beta-blockers Use	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Yes	Placebo	63	2.43	-1.39			0.7232
	Mirabegron 50 mg	50	3.23	-1.91	-0.17	0.6382	
	Mirabegron 100 mg	55	2.53	-1.38	0.06	0.8759	
	Tolterodine SR 4 mg	57	2.89	-1.46	0.13	0.7109	
No	Placebo	228	2.74	-1.06			
	Mirabegron 50 mg	243	2.75	-1.56	-0.48	0.0071	
	Mirabegron 100 mg	226	2.97	-1.55	-0.38	0.0348	
	Tolterodine SR 4 mg	243	2.57	-1.16	-0.18	0.3257	

Source: Statistical reviewer's calculation

Table 36: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Beta-blockers use at baseline – Study 178-CL-047

Beta-blockers Use	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Yes	Placebo	52	3.27	-0.71			0.1698
	Mirabegron 50 mg	43	2.50	-1.36	-1.01	0.0151	
	Mirabegron 100 mg	47	3.13	-1.66	-1.16	0.0002	
No	Placebo	273	2.98	-1.32			
	Mirabegron 50 mg	269	2.82	-1.45	-0.22	0.2128	
	Mirabegron 100 mg	249	2.61	-1.54	-0.40	0.0222	

Source: Statistical reviewer's calculation

Table 37: Change from Baseline to Final visit in Incontinence episodes per 24 hours by Beta-blockers use at baseline – Study 178-CL-074

Beta-blockers Use	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Yes	Placebo	39	2.59	-1.23			0.6055
	Mirabegron 25 mg	45	2.53	-1.36	-0.17	0.6980	
	Mirabegron 50 mg	34	2.43	-1.74	-0.60	0.1955	
No	Placebo	223	2.40	-0.83			
	Mirabegron 25 mg	209	2.67	-1.45	-0.47	0.0144	
	Mirabegron 50 mg	223	2.52	-1.32	-0.42	0.0244	

Source: Statistical reviewer's calculation.

Table 38: Change from Baseline to Final visit in micturitions per 24 hours by Beta-blockers use at baseline – Study 178-CL-046

Beta-blockers Use	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Yes	Placebo	83	11.41	-1.28			0.1418
	Mirabegron 50 mg	69	11.38	-1.16	0.11	0.9697	
	Mirabegron 100 mg	86	11.93	-1.80	-0.41	0.2727	
	Tolterodine SR 4 mg	86	11.87	-1.65	-0.29	0.4301	
No	Placebo	397	11.78	-1.39			
	Mirabegron 50 mg	404	11.69	-2.08	-0.71	<.0001	
	Mirabegron 100 mg	392	11.42	-1.73	-0.44	0.0099	
	Tolterodine SR 4 mg	389	11.47	-1.39	-0.24	0.1645	

Source: Statistical reviewer's calculation

Table 39: Change from Baseline to Final visit in micturitions per 24 hours by Beta-blockers use at baseline – Study 178-CL-047

Beta-blockers Use	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Yes	Placebo	63	11.27	-0.54			0.6146
	Mirabegron 50 mg	53	11.43	-1.64	-1.08	0.0352	
	Mirabegron 100 mg	70	11.58	-1.60	-0.96	0.0452	
No	Placebo	370	11.55	-1.08			
	Mirabegron 50 mg	372	11.85	-1.72	-0.54	0.0076	
	Mirabegron 100 mg	342	11.67	-1.78	-0.66	0.0014	

Source: Statistical reviewer's calculation

Table 40: Change from Baseline to Final visit in micturitions per 24 hours by Beta-blockers use at baseline – Study 178-CL-074

Beta-blockers Use	Treatment	# of patients	Baseline Mean	Mean Change	LS Mean change trt vs. placebo	P-value trt vs. placebo	P-value for interaction term
Yes	Placebo	60	10.73	-0.90			0.8293
	Mirabegron 25 mg	73	11.61	-1.81	-0.60	0.1759	
	Mirabegron 50 mg	63	11.30	-1.73	-0.67	0.1409	
No	Placebo	355	11.61	-1.20			
	Mirabegron 25 mg	337	11.69	-1.63	-0.43	0.0258	
	Mirabegron 50 mg	363	11.73	-1.60	-0.37	0.0498	

Source: Statistical reviewer's calculation.

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

/s/

JIA GUO
06/07/2012

MAHBOOB SOBHAN
06/07/2012

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 20-2611

Applicant: Astellas Pharma Global Development INC.

Stamp Date: Aug 29, 2011

Drug Name: Mirabegron

NDA/BLA Type: New

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

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	√			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	√			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated.	√			
4	Data sets in EDR are accessible and conform to applicable guidances (e.g., existence of define.pdf file for data sets).	√			The size of data sets for patient diary data, labs and vital signs is huge. Having difficulty to open them in software.

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? Yes

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

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

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

File name: Statistics Filing Checklist for a New NDA_202611

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

There is no review issue noted at this time.

Requests to the Applicant on 74-day letter:

- Submit statistical analysis programming code of primary and key secondary efficacy endpoints for study 178-cl-046, 047, and 074.

Jia Guo, Ph.D.	10/27/2011
_____ Reviewing Statistician	_____ Date
Mahboob Sobhan, Ph.D.	10/27/2011
_____ Supervisor/Team Leader	_____ Date

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

/s/

JIA GUO
10/27/2011

MAHBOOB SOBHAN
10/27/2011



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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

IND Number: 69,416 / Serial 0000

Drug Name: Compound YM178
[REDACTED] (b) (4)
mirabegron

Indication: Treatment of overactive bladder

Applicant: Astellas Pharma Global Development, Inc.
Deerfield, Illinois

Testing Facilities: [REDACTED] (b) (4)

Date: Submitted 30 December 2009
To reviewer 6 February 2010
Final data submitted 8 July 2010

Review Priority: Standard

Biometrics Division: Division 6

Statistical Reviewer: Steve Thomson

Concurring Reviewer: Team Leader: Karl Lin, Ph. D.

Medical Division: Reproductive and Urologic Products

Toxicologist: Eric Andreasen, Ph.D.
Lynnda Reid, Ph.D.

Project Manager: Meredith Alpert
Eufrecina Deguia

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

Table of Contents

1. EXECUTIVE SUMMARY.....	3
1.1. CONCLUSIONS AND RECOMMENDATIONS	3
1.2. BRIEF OVERVIEW OF THE STUDIES	6
1.3. STATISTICAL ISSUES AND FINDINGS	6
1.3.1. <i>Statistical Issues</i>	6
1.3.2. <i>Statistical Findings</i>	10
2. INTRODUCTION.....	10
2.1. OVERVIEW.....	10
2.2. DATA SOURCES	10
3. STATISTICAL EVALUATION	11
3.1. EVALUATION OF EFFICACY.....	11
3.2. EVALUATION OF SAFETY	11
3.2.1. PROTOCOL 7675-102: 104-WEEK ORAL GAVAGE CARCINOGENICITY STUDY WITH YM178 IN RATS.....	11
3.2.2. CTL NUMBER 178-TX-031: 104-WEEK ORAL GAVAGE CARCINOGENICITY STUDY WITH YM178 IN MICE	16
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS.....	20
5. SUMMARY AND CONCLUSIONS.....	20
5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	20
5.2. CONCLUSIONS AND RECOMMENDATIONS	20
APPENDICES:.....	21
APPENDIX 1. FDA SURVIVAL ANALYSIS.....	21
APPENDIX 2. FDA NONPARAMETRIC BAYESIAN SURVIVAL ANALYSIS.....	25
APPENDIX 3. FDA POLY-K TUMORIGENICITY ANALYSIS	30
APPENDIX 4. REFERENCES.....	36

1. EXECUTIVE SUMMARY

Using the standard boilerplate language, the Sponsor's reports indicate that the objective of these rat and mouse studies was to investigate the carcinogenic potential and provide samples for toxicokinetics of the of the test article, YM178, when administered daily via oral gavage to rats and mice for at least 104 weeks

1.1. Conclusions and Recommendations

This submission summarizes the results of a two year rat study and a two year mouse study to assess the carcinogenic potential of test compound YM178 in rats and mice when administered daily via oral gavage. Gross aspects of the study designs are summarized in the following tables, for each gender in each species:

Table 1. Design of Rat Study

Treatment Groups	# animals per study per gender	Male Rats		Female Rats	
		Dosage (mg/kg/day)	Concentration %	Dosage (mg/kg/day)	Concentration %
1. Vehicle	60	0	0	0	0
2. Low	60	12.5	2.5	25	5
3. Medium	60	25	5	50	10
4. High	60	50	10	100	20

Table 2. Design of Mouse Study

Treatment Groups	# animals per study per gender	Dosage (mg/kg/day)	Concentration %
1. Vehicle	70	0	0
2. Low	70	25	5
3. Medium	70	50	10
4. High	70	100	20

Somewhat more detailed descriptions of the studies are provided in Sections 3.2.1 and 3.2.2, below. Note that the vehicle groups are sometimes referred to as "control groups." The low, medium, and high dose groups, groups 2 to 4, are referred to as "actual dose groups" or "treated groups" while all four groups are referred to as "study groups."

The following table, Table 3, summarizes the tests comparing survival profiles across study groups in the tumorigenicity data sets:

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1-4	0.1255	0.1453	0.0001	<0.0001
No trend over Groups 1-4	0.0481	0.0527	<0.0001	<0.0001
No Difference Between Groups 1 vs 4	0.0408	0.0453	<0.0001	<0.0001
Mice Homogeneity over Groups 1-4	0.6391	0.5734	0.4393	0.4348
No trend over Groups 1-4	0.5444	0.4937	0.5110	0.5038
No Difference Between Groups 1 vs 4	0.4437	0.3415	0.2846	0.2926

Figures A.1.1 through A.1.4, in Appendix 1, provide survival curves for each of the four species by gender combinations. From Figure A.1.2, in female rats there is a strong tendency for increasing mortality over increasing dose. This is quite consistent with the various tests of no trend, homogeneity overall, and no pairwise differences between the high dose and control (all six $p \leq 0.0001$). In figure A.1.1 for male rats there seems to be a slight decreasing trend in mortality over increasing doses. This is consistent with the marginally statistically significant tests of trend over all dose groups (Logrank $p = 0.0481$, Wilcoxon $p = 0.0527$). Similarly, this explains the marginal statistical significance of the test of differences in mortality between the high dose and vehicle groups (Logrank $p = 0.0408$, Wilcoxon $p = 0.0453$). However the tests of no overall homogeneity over all study groups are not statistically significant at the usual 0.05 level (Logrank $p = 0.1255$, Wilcoxon $p = 0.1453$).

In both genders in mice, the product limit estimates of the survival seem to suggest a slight, rather vague tendency for the increasing mortality over increasing dose. However, none of the tests of trend or homogeneity over study groups, or pairwise differences between the high dose and vehicle are statistically significant are statistically significant at the usual 0.05 level (Males: all six $p \geq 0.3415$, Females: all six $p \geq 0.2846$).

An experimental Bayesian nonparametric analysis of survival is given in Appendix 2. A Bayesian analysis postulates that probability is a useful measure of uncertainty about the parameters of a statistical model. The analysis will assess the impact of the data on this prior uncertainty. For male rats this analysis indicates that the rough posterior probability that the trend over study groups is decreasing (i.e., slope less than 0) is about 0.40-0.45. So the probability that the survival time increases over dose, is roughly 0.55-0.60. In female rats it seems that the the probability that mean survival in the high dose group is less than control is roughly about 0.88 to 0.95, suggestive of a dose effect on survival. In male mice, since 0 is near the middle of each highest posterior density (HPD) interval, there is no evidence of strong differences between the actual treatment groups and the vehicle group. In female mice, again using another back of the envelope computation, the probability that mean survival in the each of the three actual dose groups is less than control is very roughly about 0.80-0.90 or so in each group.

The significance levels of the tests of tumorigenicity in the FDA analysis are based on poly-k tests applied to the data sets provided by the Sponsor. The poly-k test modifies the original Cochran-Armitage test of dose related trend over study groups in the occurrence of an event to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). One problem with any such tumorigenicity analyses is that for each tumor-organ-gender-study combination there is one test of significance for each comparison of an actual treatment group to controls plus a test of overall trend. This implies a large number of tests, necessitating a multiplicity adjustment. For two species, two gender per species studies the so-called Haseman-Lin-Rahman rules adjust for this multiplicity of tests of tumorigenicity by modifying the interpretation of the usual significance level (i.e. “p-value”). These rules specify that, in a standard two-species, two gender submission, for tests of trend at a roughly overall 0.10 (10%) false positive error rate (i.e. Type I error), one might claim statistical significance if the observed significance level is 0.025 for rare tumors (with a historical control incidence less than 1%) and 0.005 (incidence at or greater than 1%) for common tumors. Similarly, tests comparing the high dose group to controls would be considered statistically significant if the observed significance level is 0.05 for rare tumors and 0.01 for common tumors. This adjustment for multiplicity is discussed in Section 1.3.1.5 below. Table 4 below, displays the tumor incidence in both rats and mice, respectively, as well as the results of tests of no differences between treatments for those neoplasms that had at least one test that achieved a nominal 0.05 level of significance. Complete tables are provided in Appendix 3.

Table 4. Potentially Statistically Significant Neoplasms in Rats and Mice

	Incidence				Significance Levels					
	Veh	Low	Med	Hi	Trend	Trend 0-2	High vs Veh	Medium vs Veh	Low vs Veh	
Male Rats										
Thyroid										
B-Adenoma, Follicular Cell	0	0	3	2	0.0879	0.0385	0.2675	0.1364	.	
Female Rats										
Skin										
M-Sarcoma with bone formation	0	0	0	2	0.0302	.	0.1366	.	.	
Female Mice										
LIVER										
Hepatocellular adenoma	1	11	7	3	0.5876	0.0468	0.2922	0.0280	0.0017	

In Table 4 above, in female rats, following the adjustment for multiplicity to get an overall rough 10% error rate and using the incidence in the vehicle group to decide if a tumor is rare or not, we would conclude, that the test of malignant sarcoma with bone formation was not quite statistically significant ($p = 0.0302 > 0.025$), though close. If one accepts the further inflation of Type I error by the other tests, we would similarly conclude, that the test of trend over study groups in benign follicular adenoma in male rats was also not quite statistically significant ($p = 0.0385 > 0.025$). Using the incidence in the vehicle group to determine whether a tumor was rare or common, one would classify hepatocellular adenoma of the liver as a common tumor. Then, for this tumor in female mice the test of dose related trend over the first three groups was not statistically significant ($p = 0.0468 > 0.005$), as would be the pairwise comparison of the medium dose group to control ($p = 0.0280 > 0.01$). The only potentially

statistically significant comparison is between the low dose and vehicle in hepatocellular adenoma of the liver in female mice ($p = 0.0017 < 0.01$) and that would come at the cost of inflation of the type I error to some level above 10%. In both rats and mice, no other test of trend or test of pairwise differences between the high dose and vehicle was statistically significant at a 0.05 level, let alone at one of the multiplicity adjusted significance levels.

1.2. Brief Overview of the Studies

This submission had a standard rat study:

^{(b) (4)} **Study Number 7675-102: 104-Week Oral Gavage Carcinogenicity Study with YM178 in Rats,**

plus a standard mouse study:

^{(b) (4)} **Number 178-TX-031: 104-Week Oral Gavage Carcinogenicity Study with YM178 in Mice,**

to assess the carcinogenic potential of compound YM178 in rodents. Each gender in each study involved four treatment groups including a vehicle control, with 60 animals per group in the rat study and 70 per group in the mouse study.

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. Survival Analysis:

The survival analyses presented here are based on both the log rank test and the Wilcoxon test comparing survival curves. The log rank tests tend to 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 log rank 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. Both tests were used to test both homogeneity of survival among the treatment groups and the effect of dose on trend in survival. Appendix 1 reviews the specific animal survival analyses in more detail. The results of the similar Sponsor's analyses are summarized in Sections 3.2.1.1 and 3.2.2.1.

An experimental Bayesian nonparametric analysis of survival is given in Appendix 2. A Bayesian analysis takes preliminary estimates of the probability that a parameter in the statistical model of the process being analyzed satisfies certain criteria, and uses the data to

modify these probability estimates. In this context it answers questions about what is the probability that the parameter satisfies the criteria.

1.3.1.2. Multiplicity of Tests on Survival:

Using the logrank and Wilcoxon tests, there are six tests of survival in each species by gender combination. If we were to assume the tests are independent across comparisons, which clearly they are not, and assume that there is absolutely no difference in survival, the probability of at least one statistically significant result in each species, at the usual 0.05 level, is about 0.265, and about 0.460 of at least one statistically significant result in at least one gender in at least one study. Such is the possible price paid for the multiplicity of hypothesis tests in the frequentist paradigm. Note that the Bayesian test is based on a hierarchical model and this provides an inherent adjustment for the multiplicity of tests.

1.3.1.3. Tests on Neoplasms:

The Sponsor's rat report indicates that incidental tumors were analyzed using logistic regression techniques, while the fatal or mortality independent tumors were analyzed by time to event techniques. The report indicates that these were combined using methods similar to a typical Peto analysis. The mouse report indicates that tumors were analyzed using Peto techniques. Note that these tests require accurate determination of whether a tumor is fatal or incidental.

Appendix 3 presents the results from the FDA poly-k analysis on tumor incidence in rats and mice. The poly-k test is a modification of the original Cochran-Armitage test of trend in response to dose, adjusted for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that the poly-k modification of the Cochran-Armitage tests of trend has been recommended over the corresponding Peto tests.

1.3.1.4. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms necessitates a number of statistical tests, which in turn requires an adjustment in experiment-wise Type I error (i.e., the probability of rejecting a true null hypothesis). Based on his extensive experience with such carcinogenicity analyses in standard laboratory rodents, for pairwise tests between the high dose group and controls in two species, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate in standard two species, two gender studies, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Similarly, Lin and Rahman (1998) showed that tests of trend over all doses should be tested at a 0.025 (2.5%) level for rare tumors and 0.005 (0.5%) for common tumors. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorigenicity when there actually is such a relation).

Note that significance levels of the pairwise tests between the vehicle dose group 1 and each of the three actual dose groups are also provided. Including these tests will increase the

overall type I error rate. Even if one uses the Haseman-Lin-Rahman rules, the overall type I error associated with including the tests between the vehicle and the low and medium dose groups may be considerably larger than the rough 10% when these rules are restricted to the test of trend and pairwise differences between the high dose and vehicle.

1.3.1.5. Validity of the Designs:

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

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

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

Lin and Ali (2006), quoting work by Haseman, have suggested that in standard laboratory rodent species, a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. Note that as a percentage of animals that survived to week 91, this criterion is met in rats in all dose groups in both genders, and is clearly well exceeded in both mouse genders. (Please see tables 8 and 9 on page 14, tables 14 and 15 on pages 18 and 19).

The mean weight values in the following tables were taken from the Sponsor's rat and mice reports (Tables 5 and 6, Rat study pages 128-139; Table 7, Mice study pages 230-260). The change from baseline is the simple difference between means and is not mortality adjusted.

Table 5. Mean Weights for Mice and Rats (in g)

Rats Dose Group	Males				Females			
	Week		Change from baseline	% change relative to vehicle	Week		Change from baseline	% change relative to vehicle
	1	105			1	105		
Vehicle	118	382	264		98	254	156	
Low ¹	119	373	254	96.2%	97	254	157	100.6%
Medium	118	364	246	93.2%	97	249	152	97.4%
High	118	338	220	86.6%	97	245	148	94.9%

Mice Dose Group	Males				Females			
	Week		Change from baseline	% change relative to vehicle	Week		Change from Baseline	% change relative to vehicle
	1	105			1	105		
Vehicle	21.1	39.2	18.1		18.3	40.0	21.7	
Low ¹	21.0	37.6	16.6	91.7%	18.2	35.2	17.0	78.3% ¹
Medium	21.0	35.6	14.6	80.7%	18.0	33.7	15.7	72.4%
High	20.8	35.1	14.3	79.0%	18.0	32.9	14.9	68.7%

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag *et al* (1976) recommend that the MTD “is taken as ‘the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal’s natural life span’ ” From Table 4, above, in rats this criterion is nearly exceeded in the low dose in males and is clearly exceeded by each dose in female mice, and in the high and medium doses in male mice. This may be evidence that the MTD was exceeded in these dose groups.

The Sponsor’s rat report notes that “Mean food consumption values were generally higher (most often significantly higher) than that of controls for males and females given YM178. The mean food consumption values were generally increased in a dose-related manner for males. Females given 100 mg/kg/day generally had the greatest mean food consumption values, while females given 25 or 50 mg/kg/day generally had similar values.” (page 26 of rat report)

The Sponsor’s mice report notes that “Food consumption was statistically significantly increased throughout the dosing period in a dose-related manner in all YM178-treated groups compared with control . . . The maximum difference control was week 3 for males (144%) and week 10 for females (135%).” (page 21 of mice 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. If dosing is close to the MTD one would expect slightly higher mortality due to toxicity, but not so much that it largely reduces the number of animals exposed to the drug. In male rats there is evidence of lower mortality in the high dose group compared to the other groups. In female rats all actual dose groups seem to have similar mortality, somewhat higher than the vehicle. In both mice genders, but particularly in females, mortality in the high dose group does seem to differ from the control groups. A related way to assess whether or not the MTD was achieved is to measure mortality not associated with any identified tumor. Table 5, below, indicates that the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors (i.e., the “Event”):

Table 6. Natural Death with No Identified Tumor (Male/Female)

Rats		1. Vehicle 0 mg/kg	2. Low 12.5/25 ¹ mg/kg	3. Medium 25/50 ¹ Mg/kg	4. High 50/100 ¹ mg/kg
Male	Event	1	1	3	2
	No event	59	59	57	58
Female	Event	4	9	13	27
	No event	56	51	47	33

¹The first amount is the male dose, the second is the female dose.

Table 6. (cont.) Natural Death with No Identified Tumor (Male/Female)

Mice		1. Vehicle 0 mg/kg	2. Low 25 mg/kg	3. Medium 50 mg/kg	4. High 100 mg/kg
Male	Event	0	5	4	7
	No event	70	65	66	63
Female	Event	1	2	5	2
	No event	69	68	65	68

The apparent simple frequency results above for male rats are consistent with the tests comparing the survival curves (Males: Logrank $p = 0.5748$, Wilcoxon $p = 0.5971$). That is, there is no evidence of difference in deaths without tumor when comparing the high dose in male rats to the vehicle. However, in female rats both the simple incidence table and the tests comparing the survival curves show clear evidence of differences between the high dose and vehicle in the event (Females: both Logrank and Wilcoxon $p \leq 0.00005$). Results are reversed in mice. In male mice, results in the simple incidence table suggest, and the tests comparing the survival curves confirm evidence of differences in the early death without tumor (Males: Logrank $p = 0.0069$, Wilcoxon $p = 0.0070$) while both the incidence table and the tests in females comparing the high dose to vehicle in females suggest no difference (Females: Logrank $p = 0.5485$, Wilcoxon $p = 0.5395$). Again, while, like the other observations above, these require the expertise of the toxicologist, they may be evidence that the MTD was exceeded in female rats.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

This submission summarizes the results of a two year rat study and a two year mouse study to assess the carcinogenic potential of compound YM 178 by daily oral gavage.

2.2. Data Sources

Each hidden within several levels of DARRTS, the Sponsor provided a SAS transport data sets for both studies labeled tumor.sas7bdat. Both data sets required extensive preparation for analysis, since numerous null records were included, and variable names and characteristics differed from those requested by the FDA Office of Biostatistics.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1. Protocol 7675-102: 104-Week Oral gavage Carcinogenicity Study with YM178 in Rats,

STUDY DURATION: 105 Weeks

EXPERIMENTAL START DATE: 17 August 2005

EXPERIMENTAL END DATE: 22 August 2007 (date of last necropsy)

RAT STRAIN: (b) (4) CDF[®]F-344/DuCrI Rats

ROUTE: Oral gavage

The basic design of the rat study has four dose groups, summarized in Table 7, actually a repeat of Table 1, below:

Table 7. Design of Rat Study (Dose Volume 5 mL/kg)

Treatment Groups	# animals per study per gender	Male Rats		Female Rats	
		Dosage (mg/kg/day)	Concentration %	Dosage (mg/kg/day)	Concentration %
1. Vehicle	60	0	0	0	0
2. Low	60	12.5	2.5	25	5
3. Medium	60	25	5	50	10
4. High	60	50	10	100	20

In addition, 4 animals of each gender were used for the toxicokinetic study control group and 12 of each gender for the corresponding toxicokinetic actual dosing groups (groups 5-8, respectively). The Sponsor states that animals were randomly allocated to treatment, stratified by weight. Animals were dosed once daily in the morning, seven days per week. The main study animals were dosed for at least 104 weeks, while the toxicokinetic animals were treated for at least 52 weeks.

Animals were housed individually in stainless steel cages. Food and water were available *ad libitum*, except at the time of necropsy.

The Sponsor justifies dosing levels as follows: "A 13-week oral gavage toxicity study with YM178 was conducted in F344 Rats (10 animals/sex/group) at dose levels 10, 30, 100, and 300 mg/kg/day. As a result, deaths of two female animals in the 300-mg/kg/day group at

Weeks 10 and 13 were reported. Also, remarkable inhibition for body weight increase was observed for both sexes in the same dose group.

“In a 26-week oral gavage toxicity study with YM178 conducted in F344 Rats (12 animals/sex/group) at dose levels 10, 30, and 100 mg/kg/day, food and water intake was increased for animals given 10 mg/kg/day or greater. There was a 17% inhibition of body weight increase compared with control group values, increased alanine aminotransferase and aspartate alanine aminotransferase in the 30-mg/kg/day group. Increasing ratios of liver weight against body weight and acidophilic changes were observed for both sexes at doses of 30 mg/kg/day or greater. At 100 mg/kg/day, there was a 36% inhibition of body weight increase for male animals reported, while no such changes were observed with female animals.

“From the fact that inhibition of body weight increase was observed with 10% of male animals at 30 mg/kg/day, the sponsor has set 50 mg/kg/day as highest dose, and with common ratio of 2, selected 25 and 12.5 mg/kg/day for lower doses. For female animals, they consider 100 mg/kg/day, has no effect on survival rate and set 100 mg/kg/day highest dose for the study, and with common ratio of 2 to select 50 and 25 mg/kg/day for greater lower doses.” (pages 15 and 16 of report).

3.2.1.1. Sponsor’s Results and Conclusions

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

Survival analysis:

The Sponsor’s statistical report summarizes survival results as “in the males, a significant negative trend in mortality was noted (Cox-Tarone: $0.0332 \leq p \leq 0.0359$; Gehan-Breslow: $p = 0.0358$) with a significant decrease in the animals given 25 or 50 mg of YM178/kg of body weight/day (mg/kg/day) when compared with animals given 0 mg/kg/day (Cox-Tarone: $p = 0.0229$ and 0.0375 , respectively; Gehan-Breslow: $p = 0.0220$ and 0.0325 , respectively). ... [I]n the females, a significant positive trend in mortality was noted (Cox-Tarone: $p=0.0000$; Gehan-Breslow: $p = 0.0000$) with a significant increase in the animals given 100 mg/kg/day when compared with animals given 0 mg/kg/day (Cox-Tarone: $p = 0.0001$; Gehan-Breslow: $p = 0.0001$).” (page 57 of report).

Tumorigenicity analysis:

The Sponsor’s statistical report also summarizes carcinogenicity results as follows: “in the males, a statistically significant negative trend in the pituitary adenoma incidences was noted ($p = 0.0000$), with significant decreases in the incidences for the animals given 12.5, 25, or 50 mg/kg/day when compared with animals given 0 mg/kg/day ($p = 0.0109$, 0.0000 , 0.0000 , respectively). Also in the males, a significant positive trend was noted for squamous cell papilloma in the skin ($0.0347 \leq p \leq 0.0576$). No significant increase was noted in any of the other dose levels when compared with animals given 0 mg/kg/day. This apparent positive trend is not due to a dose-response because the animals given the 12.5 or 25 mg/kg/day, along with

animals given 0 mg/kg/day, exhibited no incidence. Only the animals given 50 mg/kg/day in this case showed any tumors (2/60). No other significant effects were noted in the male neoplastic lesions.” (page 57 of report)

The report continues: “in the females, the pituitary adenoma combined with carcinoma incidences in the animals given 25 or 100 mg/kg/day showed significant decreases when compared with animals given 0 mg/kg/day ($p = 0.0143$ and 0.0249 , respectively) with no significant corresponding trend ($p = 0.0688$). A significant positive trend was noted for malignant sarcoma with bone formation in the skin ($0.0076 \leq p \leq 0.0152$). The animals given 25 or 50 mg/kg/day did not show any lesion in this case, and the incidence rate for animals given 100 mg/kg/day (2/60) was not significantly increased ($p = 0.1141$) when compared with animals given 0 mg/kg/day. Also in the females, endometrial stromal polyp in the uterus and cervix combined showed a marginally significant increase in the animals given 25 mg/kg/day [13/60 ($p=0.0486$)]. The trend in this case actually negative because the incidence rate for animals given 100 mg/kg/day (3/60) was smaller than the incidence rate for animals given 0 mg/kg/day (7/60). This apparent increase for the animals given 25 mg/kg/day is probably due to background noise in this lesion. No other significant effects were noted in the female neoplastic lesions.” (page 58 of report)

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

The following tables (Table 8 for male rats, Table 9 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. In these tables the terminal period only includes those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period. The Kaplan-Meier survival plots in Appendix 1 provide a more detailed picture of the profile of mortality losses.

Table 8. Summary of Male Rats Survival (dosed at mg/kg/day)

Period (Weeks)	Vehicle 0	Low 12.5	Medium 25	High 50
1-52	1/60 98.3%	0/60 100%	2/60 96.7%	2/60 96.7%
53-78	5/59 90.0%	2/60 96.7%	3/58 91.7%	1/58 95.0%
79-91	9/54 75.0%	8/58 83.3%	4/55 85.0%	6/57 85.0%
92-105	17/45 46.7%	15/50 58.3%	12/51 65.0%	12/51 65.0%
Terminal 105	28	35	39	39

¹ number of deaths / number at risk² overall per cent survival to end of period.**Table 9. Summary of Female Rat Survival (dosed at mg/kg/day)**

Period (Weeks)	Vehicle 0	Low 25	Medium 50	High 100
1-52	2/60 96.7%	5/60 91.7%	11/60 81.7%	23/60 61.7%
53-78	6/58 86.7%	5/55 83.3%	6/49 61.7%	5/37 53.8%
79-91	0/52 86.7%	5/50 75.0%	1/43 70.0%	5/32 45.0%
92-105	9/52 71.7%	8/45 61.7%	6/42 60.0%	3/27 40.0%
Terminal 105	43	37	36	24

¹ number of deaths / number at risk² overall per cent survival to end of period.

Table 10 below provides the significance levels of the tests of homogeneity and trend over dose groups as proposed in Section 1.3.1.1, above.

Table 10. Statistical Significances of Tests of Homogeneity and Trend in Survival in Rats

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1-4	0.1255	0.1453	0.0001	<0.0001
No trend over Groups 1-4	0.0481	0.0527	<0.0001	<0.0001
No Difference Between Groups 1 vs 4	0.0408	0.0453	<0.0001	<0.0001

Figures A.1.1 through A.1.4 in Appendix 1 provide survival curves for each of the four species by gender combinations. From Figure A.1.2, in female rats there is a strong tendency for increasing mortality over increasing dose. This is quite consistent with the various tests of trend, differences overall, and pairwise differences between the high dose and control (all six $p \leq 0.0001$). In figure A.1.1 for male rats there seems to be a slight decreasing trend in mortality over increasing doses. This is consistent with the marginally statistically significant tests of trend over all dose groups (Logrank $p = 0.0481$, Wilcoxon $p = 0.0527$). Similarly, the test of differences in mortality between the high dose and vehicle groups (Logrank $p = 0.0408$, Wilcoxon $p = 0.0453$). However the tests of no overall homogeneity are not statistically significant at the usual 0.05 level (Logrank $p = 0.1255$, Wilcoxon $p = 0.1453$).

Tumorigenicity analysis:

As discussed in Section 1.3.1.5, for common tumors, the Haseman-Lin-Rahman rules specify that for a roughly 0.10 (10%) overall false positive error rate, overall trend should be tested at a 0.025 (2.5%) level in rare tumors and at 0.005 (0.5%) in common tumors. Pairwise tests between the high dose group and control should be tested at 0.05 (5%) level in rare tumors and at a 0.01 (1%) in common tumors. Table 11 below lists the only organ by tumor combination in rats that had at least one test of trend or pairwise comparison with a nominal significance level of 0.05.

Table 11. Potentially Statistically Significant Neoplasms in Rats

	Incidence				Significance Levels				
	Veh	Low	Med	Hi	Trend	Trend			Low
						0-2	vs Veh	vs Veh	
Male Rats									
Thyroid									
B-Adenoma, Follicular Cell	0	0	3	2	0.0879	0.0385	0.2675	0.1364	.
Female Rats									
Skin									
M-Sarcoma with bone formation	0	0	0	2	0.0302	.	0.1366	.	.

In Table 11, in female rats, following the adjustment for multiplicity to get an overall rough 10% error rate and using the incidence in the vehicle treatment group to decide if a tumor is rare or not, we would conclude, that the test of malignant sarcoma with bone formation was not quite statistically significant ($p = 0.0302 > 0.025$), though close. If one accepts the inflation of Type I error by the other tests, we would similarly conclude, that the test of trend in benign follicular cell adenoma in male rats was also not quite statistically significant ($p = 0.0385 > 0.025$). No other test of trend or test of pairwise differences between the high dose and vehicle in rats was statistically significant at a 0.05 level, let alone at one of the multiplicity adjusted significance levels.

3.2.2. (b) (4) Number 178-TX-031: 104-Week Oral gavage Carcinogenicity Study with YM178 in Mice

STUDY DURATION: 105 Weeks
 EXPERIMENTAL START DATE: 30 September 2005
 EXPERIMENTAL END DATE: 1 October 2007
 MOUSE STRAIN: (b) (4) B6C3F1 Mice
 ROUTE: Oral gavage

The basic design of the mouse study has four dose groups per gender, summarized in Table 12 below (same as Table 2 above):

Table 12. Design of Mice Study (dose volume: 1 ml/100g/dose)

Treatment Groups	# animals per study per gender	Dosage (mg/kg/day)	Concentration %
1. Vehicle	70	0	0
2. Low	70	25	5
3. Medium	70	50	10
4. High	70	100	20

The Sponsor states that animals were randomly allocated to treatment, stratified by weight. Animals were housed individually in solid bottom plastic cages, with food and water available *ad libitum*.

The Sponsor justifies dosing as follows: “In the 2-week dose range finding study (Okazaki, 2003), YM178 was given at dose levels of 0, 30, 100, and 300 mg/kg/day to B6C3F1 mice. At 300 mg/kg/day, four out of 10 males and two out of 10 females died on day 1. Clonic convulsion and prone/lateral position were observed in the mice that died. At 100 mg/kg/day, the only noteworthy finding was decreased locomotor activity, which was observed in both sexes. Therefore, 200 mg/kg/day was considered appropriate for the high dose in the 13-week study.

“In the 13-week dose range finding study (Okazaki, 2003), YM178 was given at dose levels of 0, 50, 100, and 200 mg/kg/day to B6C3F1 mice. As a result, at 200 mg/kg/day, decreased locomotor activity was observed on day 1 in both sexes and one out of 12 females died on day 43 (week 7) in the toxicity group. In addition, two of 35 males in the toxicokinetic group dosed at 200 mg/kg/day died on day 1. In the surviving mice, prone position was noted from week 9 to week 13, and food consumption increased significantly throughout the dosing period in all YM178-treated groups. At 100 mg/kg/day or more, body weight increased significantly compared with the control group throughout the dosing period.

The report concludes: “Based on the results of these toxicity studies, 200 mg/kg/day was considered the lethal dose. Therefore, a dose level of 100 mg/kg/day was set as a highest dose level for both sexes, with 50 and 25 mg/kg/day serving as mid- and low-dose levels, respectively.” (page 15 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 rats.

Survival analysis:

According to the Sponsor: “There were no treatment related effects on survival . . . Numbers of mice surviving out of the original 70 per group at week 104 were as follows: ” (page 21 of report) The following table summarizes the Sponsors results on results on survival and is transcribed from the table on page 21 of the Sponsor’s report.

Table 13. Survival [actual number and as a percentage] at week 104

	Dose level of YM178 (mg/kg/day)			
	0	25	50	100
Males	61 (87%)	58 (83%)	63 (90%)	57 (81%)
Females	60 (86%)	51 (73%)	55 (79%)	55 (79%)

The discrepancy between this table and tables 14 and 15 below, seems to be due to animals that died a natural death after week 104, presumably during the terminal period.

Tumorigenicity analysis:

The Sponsor’s Table 18 in the report indicates that tests on neoplasms were Peto tests of trend, but apparently no further details are provided. The report summarizes their interpretation of results as follows: “Several tumor types occurred at high incidence, which is typical for this strain of mouse. In males these were benign adrenal gland tumors, benign and malignant liver tumors, benign lung tumors and malignant lymphomas. In females, the most frequently diagnosed tumors were malignant lymphomas, pituitary gland adenomas and to a lesser extent benign and malignant liver and lung tumors

“The only change in common tumors suggestive of a possible relationship with treatment was a reduction in lung tumors in the high dose group compare to controls in both sexes. In females benign liver tumors (hepatocellular adenomas) showed a higher incidence in the low and mid dose groups compared to the controls, but there was no relationship with dose.

“Several other tumor types occurred at low incidence in the control and treated groups. For example, there were two squamous cell carcinomas of the tongue in the control males and none in the other groups whereas there were two tubular cell adenomas of the kidney in the high dose males, one in the mid dose males, but none in the low dose or the controls and none in females. Tubular cell adenomas are uncommon in control mice, but the absence of other

effects on the kidney, their presence in this study is not considered to represent an effect of treatment.” (pages 22-23 of report)

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

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

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

Period (Weeks)	Vehicle 0	Low 25	Medium 50	High 100
1-52	0/70 ¹ 100% ²	2/70 97.1%	1/70 98.6%	2/70 97.1%
53-78	1/70 98.6%	1/68 95.7%	2/69 95.7%	3/68 92.9%
79-91	3/69 94.3%	4/67 90.0%	2/67 92.9%	2/65 90.0%
92-105	6/66 85.7%	5/63 82.9%	3/65 88.6%	6/63 81.4%
Terminal 105	60	58	62	57

¹ number of deaths / number at risk

² overall per cent survival to end of period.

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

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

Period (Weeks)	Vehicle 0	Low 25	Medium 50	High 100
1-52	0/70 ¹ 100% ²	1/70 98.6%	2/70 97.1%	2/70 97.1%
53-78	1/70 98.6%	4/69 92.9%	1/68 95.7%	2/68 94.3%
79-91	5/69 91.4%	3/65 88.6%	6/67 87.1%	4/66 88.6%
92-105	5/64 84.3%	11/62 73.9%	8/61 75.7%	8/62 77.1%
Terminal 105	59	51	53	54

¹ number of deaths / number at risk² overall per cent survival to end of period.

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

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Mice Homogeneity over Groups 1-4	0.6391	0.5734	0.4393	0.4348
No trend over Groups 1-4	0.5444	0.4937	0.5110	0.5038
No Difference Between Groups 1 vs 4	0.4437	0.3415	0.2846	0.2926

In both genders in mice, the product limit estimates of the survival seem to suggest a slight, rather vague tendency for the increasing mortality over increasing dose. However, none of the tests of trend, homogeneity, or pairwise differences between the high dose and vehicle are statistically significant at the usual 0.05 level (Males: all six $p \geq 0.3415$, Females: all six $p \geq 0.2846$).

Tumorigenicity analysis:

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

Table 17. Potentially Statistically Significant Neoplasms in Rats and Mice

	Incidence		Significance Levels						
	Veh	Low Med Hi	Trend	Trend 0-2	High vs Veh	Medium vs Veh	Low vs Veh		
Female Mice									
LIVER									
Hepatocellular adenoma	1	11	7	3	0.5876	0.0468	0.2922	0.0280	0.0017

In Table 17, in female mice, using the incidence in the vehicle group to determine whether a tumor was rare or common, one would classify hepatocellular adenoma of the liver as a common tumor. Then, for this tumor in female mice the test of dose related trend over the first three groups was not statistically significant ($p = 0.0468 > 0.005$), as would be the pairwise comparison of the medium dose group to control ($p = 0.0280 > 0.01$). The only potentially statistically significant comparison is between the low dose and vehicle in hepatocellular adenoma of the liver in female mice ($p = 0.0017 < 0.01$) and including such tests would come at the cost of inflation of the type I error to some level above 10%. No other test of trend or test of pairwise differences between the high dose and vehicle was statistically significant at a 0.05 level, let alone at one of the multiplicity adjusted significance levels.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see Section 1.1 above

APPENDICES:**Appendix 1. FDA Survival Analysis**

Simple summary life tables in mortality are presented in the report (Tables 8, 9, 14, and 15 above). Kaplan-Meier estimated survival curves across dose groups for each gender in each study are displayed in Figures A.1.1-A.1.4 below. These plots include 95% confidence intervals around each survival curve (colored area around each curve). These plots are also supported by tests of homogeneity in survival over the different treatment groups including the vehicle group, tests of trend in survival over increasing dose over all groups, and the results of pairwise comparisons between the high dose group 4 and vehicle group 1 in Table A.1.1. below. One might note that the log rank tests places greater weight on later events, while the Wilcoxon test tends to weight weights them more equally, and thus places less weight on earlier events than does the log rank test.

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1-4	0.1255	0.1453	0.0001	<0.0001
No trend over Groups 1-4	0.0481	0.0527	<0.0001	<0.0001
No Difference Between Groups 1 vs 4	0.0408	0.0453	<0.0001	<0.0001
Mice Homogeneity over Groups 1-4	0.6391	0.5734	0.4393	0.4348
No trend over Groups 1-4	0.5444	0.4937	0.5110	0.5038
No Difference Between Groups 1 vs 4	0.4437	0.3415	0.2846	0.2926

Figures A.1.1 through A.1.4, below, provide survival curves for each of the four species by gender combinations. From Figure A.1.2, in female rats there is a strong tendency for increasing mortality over increasing dose. This is quite consistent with the various tests of trend, differences overall, and pairwise differences between the high dose and control (all six $p \leq 0.0001$). In figure A.1.1 for male rats there seems to be a slight decreasing trend in mortality over increasing doses. This is consistent with the marginally statistically significant tests of trend over all dose groups (Logrank $p = 0.0481$, Wilcoxon $p = 0.0527$). Similarly, the test of differences in mortality between the high dose and vehicle groups (Logrank $p = 0.0408$, Wilcoxon $p = 0.0453$) is barely statistically significant. However the tests of no overall homogeneity over treatment groups are not statistically significant at the usual 0.05 level (Logrank $p = 0.1255$, Wilcoxon $p = 0.1453$).

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

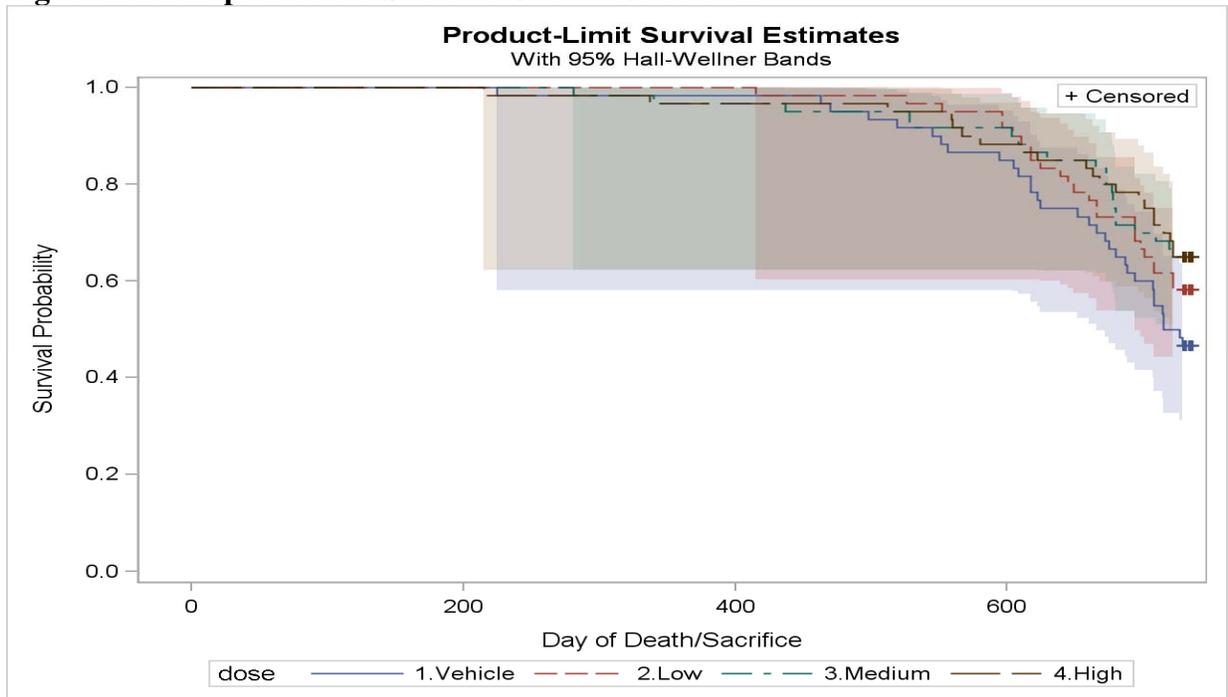
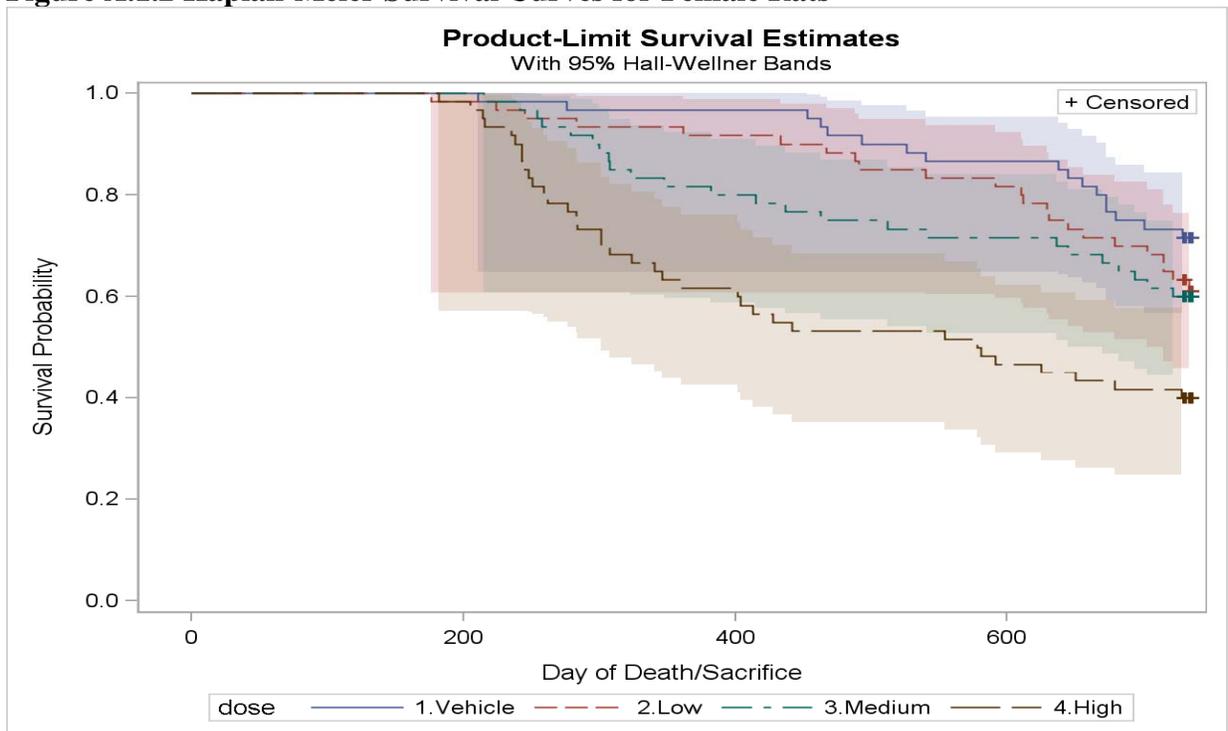


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



In both genders in mice, the product limit estimates of the survival seem to suggest a slight, rather vague tendency for the increasing mortality over increasing dose. However, none of the tests of trend, homogeneity, or pairwise differences between the high dose and vehicle are statistically significant at the usual 0.05 level (Males: all six $p \geq 0.3415$, Females: all six $p \geq 0.2846$).

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

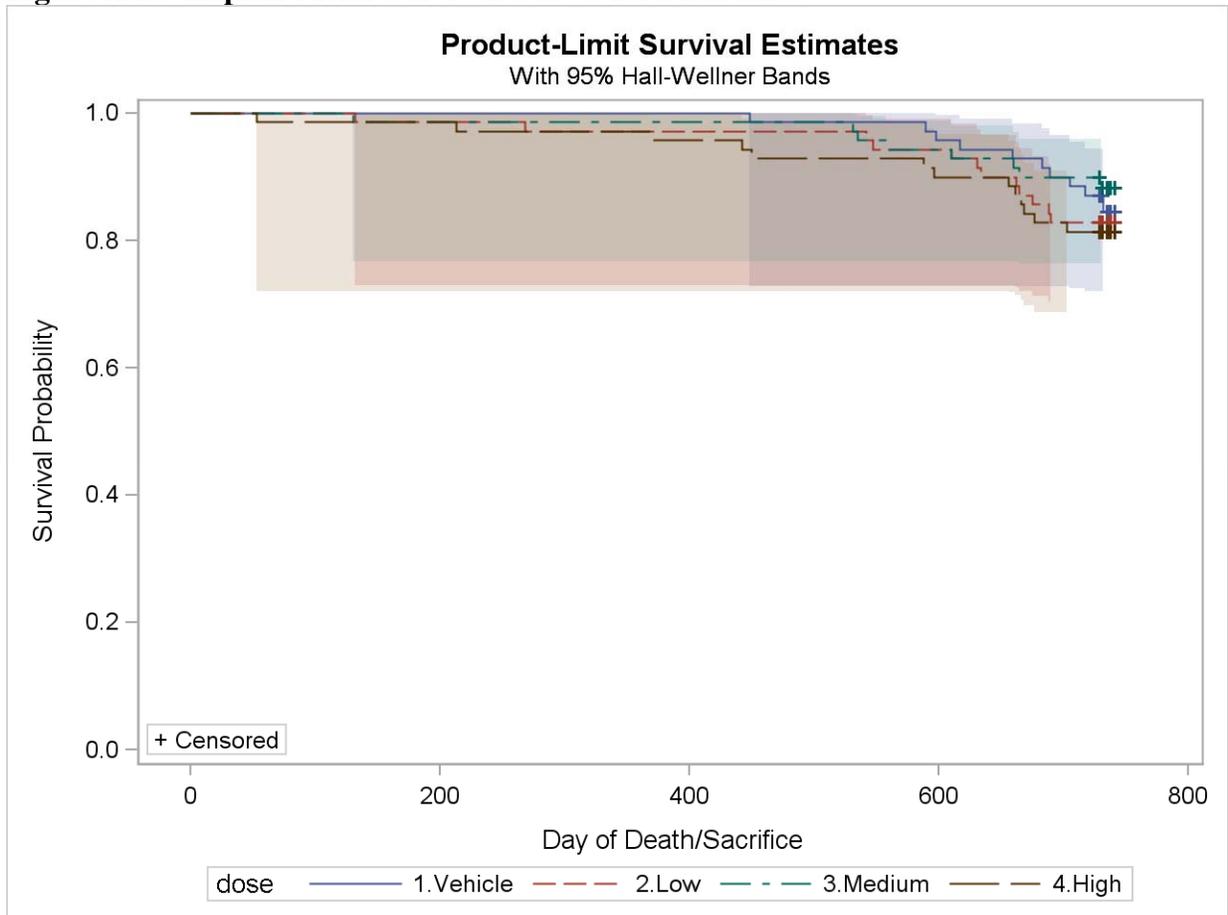
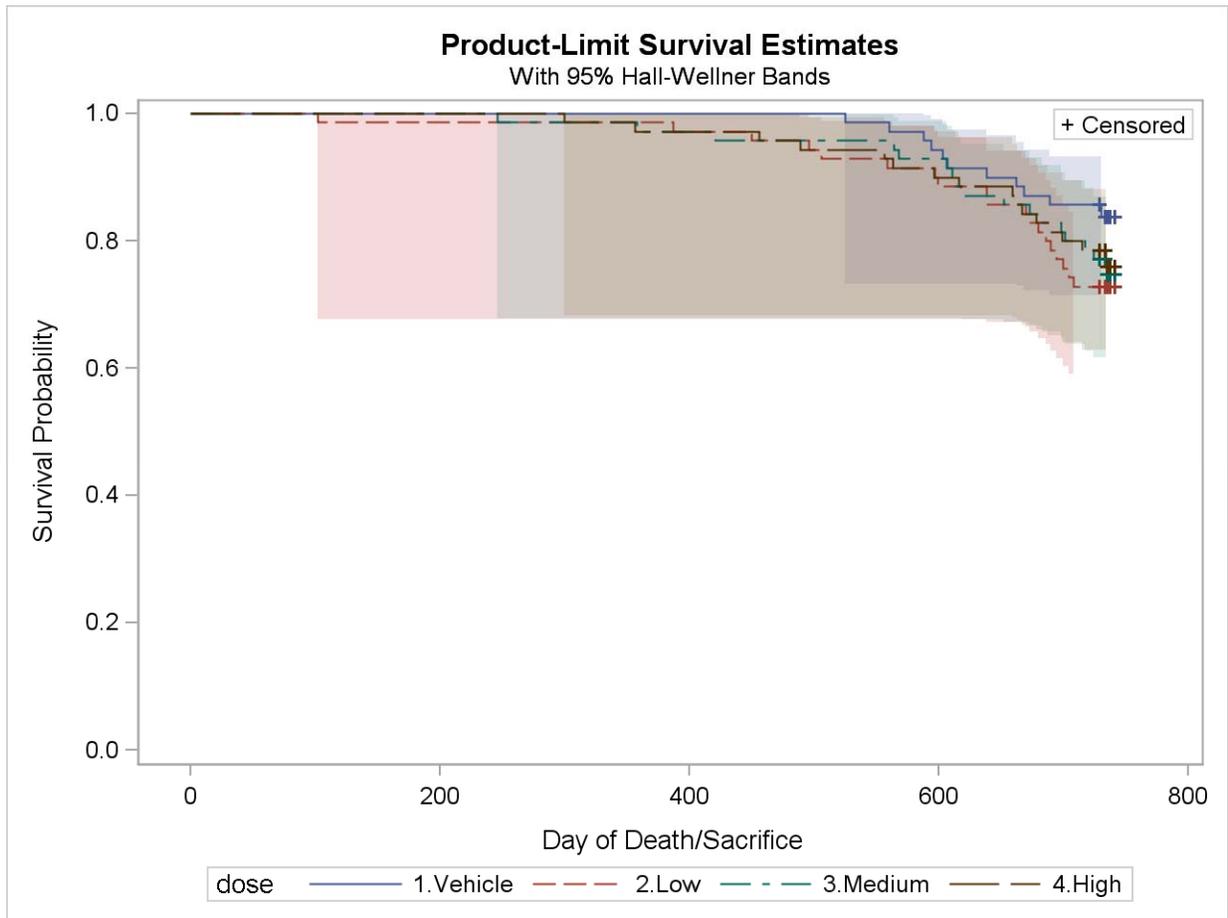


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



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. 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 some large subset of this space. The nonparametric analysis used here is based upon using a so-called Dependent Dirichlet Process (DDP) as the prior on this space of 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. For this analysis, we model the distribution of $\log(T_i)$ as a mixture of normal distributions weighted by a Dirichlet process on the normal parameters. The prior is defined as a Dirichlet process where the baseline distribution models the linear parameters as a normal distribution on the linear mean parameters and the variance parameters with a Gamma distribution. The prior of the precision parameter of the Dirichlet process is a gamma distribution. The priors for the other hyperparameters are conjugate distributions. Mathematically we can write:

$$\begin{aligned} \log(T_i) &= t_i \mid f_{X_i} \sim f_{X_i} \\ f_{X_i} &= \int N(X_i \beta, \sigma^2) G(d\beta d\sigma^2) \\ G \mid \alpha, G_0 &\sim DP(\alpha G_0) \end{aligned}$$

The distributions of the hyperparameters above are specified as follows:

$$\begin{aligned} G_0 &= N(\beta \mid \mu_b, s_b) \Gamma(\sigma^2 \mid \tau_1 / 2, \tau_2 / 2) \\ \alpha \mid a_0, b_0 &\sim \text{Gamma}(a_0, b_0) \\ \mu_b \mid m_0, S_0 &\sim N(m_0, S_0) \\ s_b \mid \nu, \Psi &\sim \text{InvWishart}(\nu, \Psi) \\ \tau_2 \mid \tau_{s1}, \tau_{s2} &\sim \text{Gamma}(\tau_{s1}, \tau_{s2}) \end{aligned}$$

This is an experimental procedure using the DPpackage (Jara, 2007) in R (R Development Core Team, 2009), and, results only should be considered as supporting. The

basic reference is de Iorio, et al (2009). The parameterization used to indicate doses was so-called dummy coding, which, in analogy with linear models as discussed in de Iorio et al (2004), implies that effect parameters for treatment doses correspond to the difference with the vehicle controls.

That is, the means mubd1, mubd2, and mubd3 below indicate the differences between the vehicle and the low, medium, and high dose groups, respectively. The HPD interval is the estimated highest posterior density interval for the parameters. Conditional on the data, the probability the indicated parameter is in the interval is 0.95.

Male Rats

Dummy Parameters	Mean	Median	Std. Dev.	95%HPD-Low	95%HPD-Upp
mub (Intercept)	7.007065	6.990593	0.454516	6.137593	7.982045
mubd1	0.804245	0.780343	0.749529	-0.628264	2.351881
mubd2	0.247170	0.235767	0.690336	-1.021069	1.730572
mubd3	0.578252	0.558068	0.784855	-0.937805	2.118423

Since 0 is near the middle of each HPD interval, there is little evidence of strong differences between the actual treatment groups and the vehicle groups. For example, using a so-called back of the envelope computation, the probability that mean survival in the high dose group is less than control (i.e., $mubd3 \leq 0$) is roughly about 0.20 to 0.25.

Slope Parameters	Mean	Median	Std. Dev.	95%HPD-Low	95%HPD-Upp
mub (Intercept)	6.790160	6.772424	0.240009	6.291727	7.233114
mubdose	0.020149	0.019519	0.111580	-0.175359	0.250332

Again, there is no strong evidence of a particular dose effect. The rough posterior probability that the slope is decreasing (i.e. less than 0) is about 0.40-0.45. So the probability that the slope is increasing, i.e., that survival time increases over dose, is roughly 0.55-0.60.

Female Rats

Dummy Parameters	Mean	Median	Std. Dev.	95%HPD-Low	95%HPD-Upp
mub (Intercept)	7.382391	7.363619	0.369658	6.660213	8.101879
mubd1	-0.265467	-0.256523	0.534596	-1.303154	0.750392
mubd2	-0.341928	-0.385477	0.597412	-1.488623	0.913975
mubd3	-0.699118	-0.716183	0.506041	-1.613145	0.331918

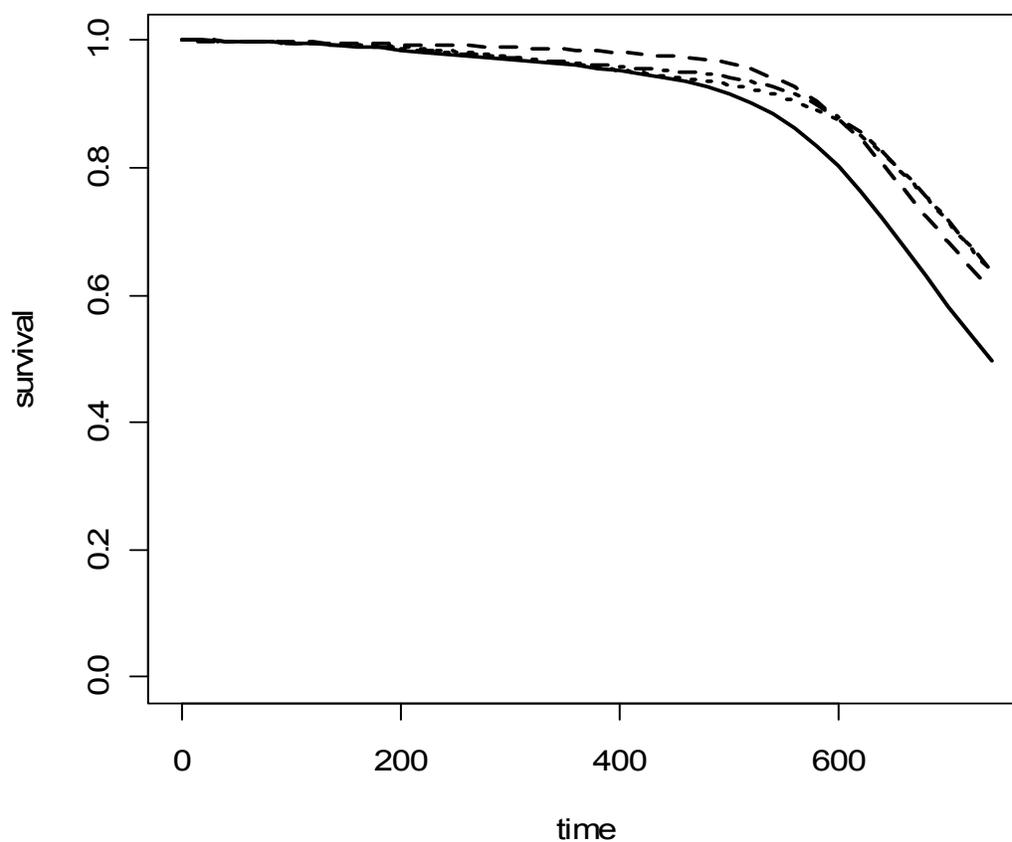
Note that 0 is in the 95% HPD interval for mubd3, corresponding to the difference between the high dose and vehicle. However, relative to the length of the interval, 0 is close to the upper boundary. Again, using a rough calculation it seems that the the probability that mean survival in the high dose group is less than control (i.e., $mubd3 \leq 0$) is roughly about 0.88 to 0.95, suggestive, but perhaps not conclusive, of a difference.

Slope Parameters	Mean	Median	Std. Dev.	95%HPD-Low	95%HPD-Upp
mub (Intercept)	7.002998	6.999322	0.201496	6.607807	7.387330
mubdose	-0.136052	-0.139317	0.089377	-0.308618	0.051633

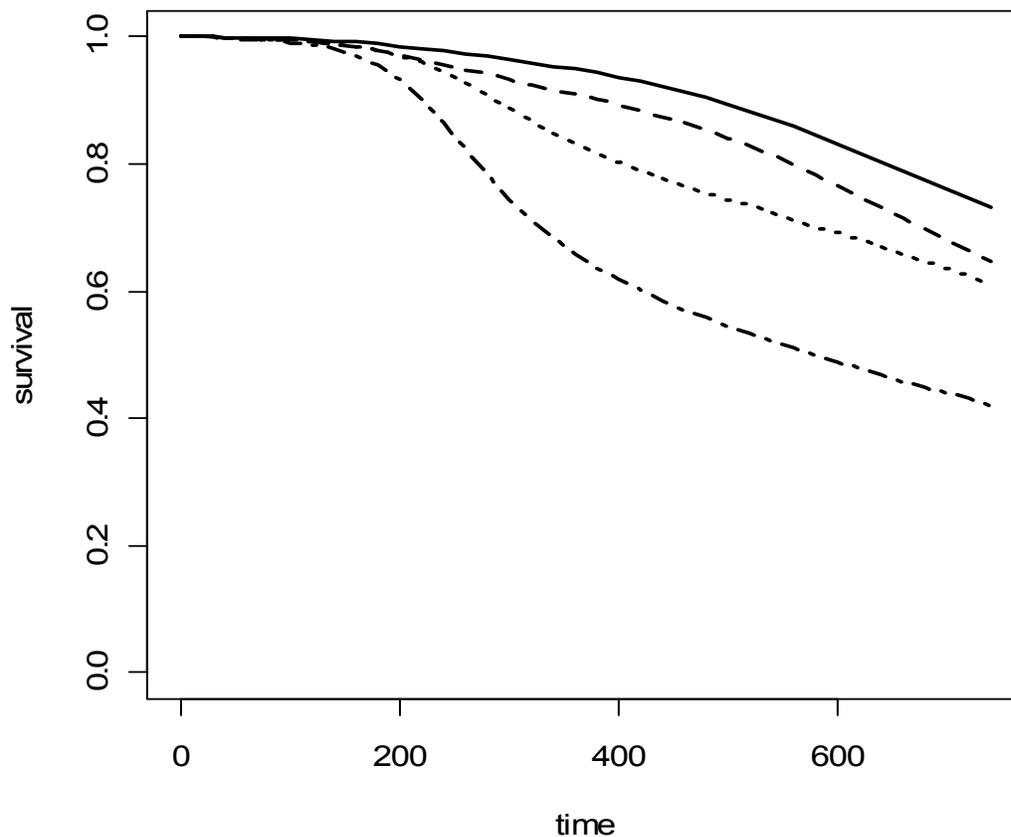
Again, while 0 is in the 95% HPD interval for dose, it is near the upper boundary, and one would estimate the probability that the difference is below 0 is roughly 0.90-0.95.

The plots below show the estimated survival curves corresponding to the four doses. The survival curve of the control dose group is drawn as a solid line, the low dose as a dashed line, the medium dose as a dotted line, and the high dose as alternating dots and dashes.

Survival Curve Male Rats:



Survival Curve Female Rats:



Male Mice

Dummy Parameters	Mean	Median	Std. Dev.	95%HPD-Low	95%HPD-Upp
mub (Intercept)	7.781888	7.756831	0.367675	7.108384	8.569024
mubd1	-0.034154	-0.052477	0.590202	-1.260329	1.125318
mubd2	0.041858	0.007660	0.589979	-1.118409	1.232268
mubd3	-0.373846	-0.379134	0.512061	-1.392800	0.576663

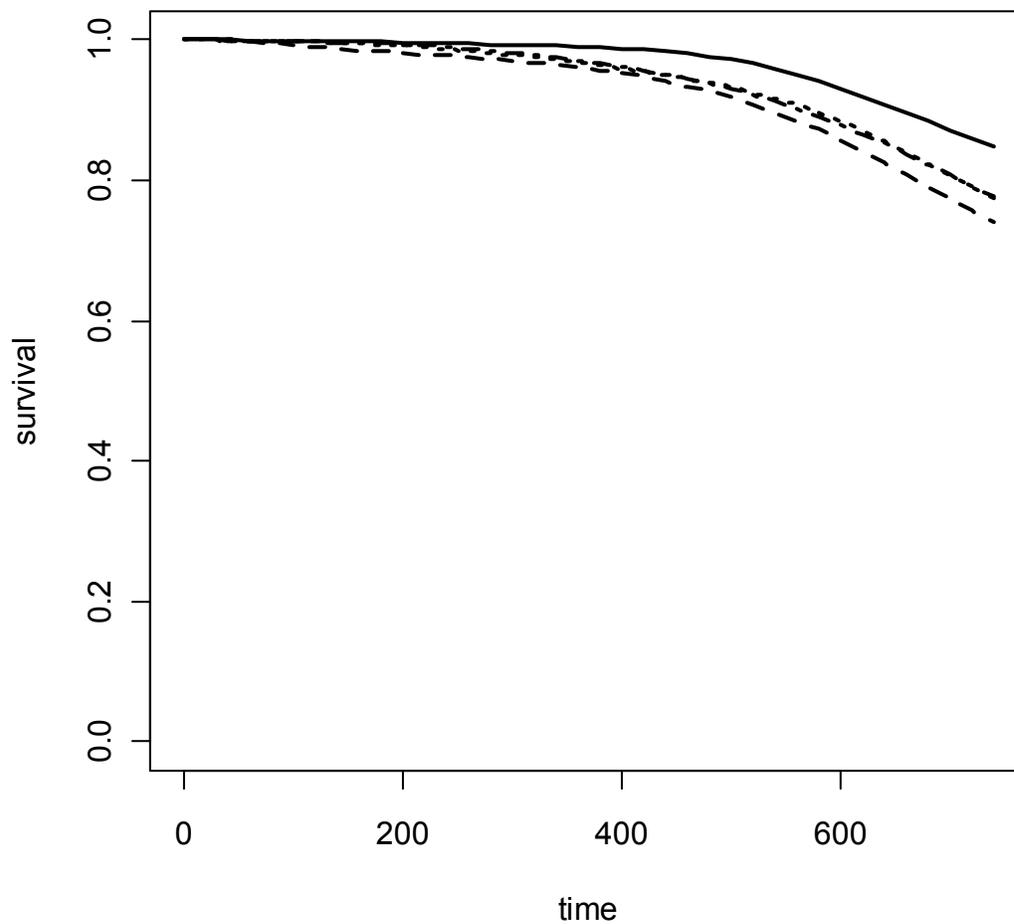
Since 0 is near the middle of each HPD interval, there is no evidence of strong differences between the actual treatment groups and the vehicle groups.

Female Mice

Dummy Parameters	Mean	Median	Std. Dev.	95%HPD-Low	95%HPD-Upp
mub(Intercept)	8.523120	8.502675	0.677043	7.217226	9.834998
mubd1	-1.172056	-1.087751	0.785125	-2.738320	0.224144
mubd2	-0.984719	-1.001306	0.757738	-2.388375	0.518479
mubd3	-1.107297	-1.076820	0.840286	-2.683829	0.578027

Since 0 is near the upper limit of each HPD interval, there is some evidence of differences between the actual treatment groups and the vehicle groups. Again, for example, using the so-called back of the envelope computation, the probability that mean survival in the each of the dose groups is less than control is roughly about 0.90 or so.

Survival Curve Female Mice:



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. These do assume all marginal totals are fixed, a debatable assumption. This assumption implies that in the pairwise tests when one dose group has no tumors of the specific type and the other does, there is only one permutation of this pattern. Since that means that the only permutation of the data is the one observed, that means that all possible permutations are as extreme as the pattern observed, and thus the significance level of the observed pattern can be logically expressed as 1.0. One could use the same sort of argument when there were no tumors of the specific type being analyzed in either column of the 2x2 table corresponding to a pairwise comparison. Then an argument could be made that the p-value for this test should also be 1.0. However, largely for readability, in the tables below these p-values are considered as missing (i.e., corresponding to a null test), denoted by “.”. Note that StatXact adjusts for the variance, which would be 0. Then the significance levels of the test statistics are based on the result of a division by 0, i.e., undefined, and hence StatXact codes these p-values as missing.

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

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.4 are often applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors (incidence > 1%) and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. As also discussed in section 1.3.1.4, using these adjustments for other tests, like the trend over the vehicle, low, and medium dose groups and the pairwise comparisons between the vehicle and the medium and low dose groups 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 through A.3.5 given below, display, in each study, for each gender, the organ by tumor combination, the number of animals with one or more of the specified tumor in each treatment group, plus the statistical significance levels of the tests of no trend over the four study groups, no trend when excluding the high dose, and no pairwise difference between the specified treatment groups and the vehicle group. Table A.3.1 in rats and mice shows the tumors that had at least one mortality adjusted test whose nominal statistical significance was at

least 0.05. Tables A.3.2 and A.3.3 display all incidences and statistical test results for male and female rats, respectively, while Tables A.3.4 and A.3.5 present similar results in male and female mice. The p-values of the poly-k test are based on exact tests from StatXact as discussed above. As also noted above, the period ‘.’ denotes the p-values of tests of dose groups with no tumors in any group. The trend test over the vehicle, low, and medium dose groups was included based on the possibility that the MTD was exceeded. Of course, determination of whether or not the MTD was exceeded requires the expertise of the toxicologist.

Table A.3.1 Potentially Statistically Significant Neoplasms in Rats and Mice

	Incidence				Significance Levels				
	Veh	Low	Med	Hi	Trend	Trend 0-2	High vs Veh	Medium vs Veh	Low vs Veh
Male Rats									
Thyroid									
B-Adenoma, Follicular Cell	0	0	3	2	0.0879	0.0385	0.2675	0.1364	.
Female Rats									
Skin									
M-Sarcoma with bone formation	0	0	0	2	0.0302	.	0.1366	.	.
Female Mice									
LIVER									
Hepatocellular adenoma	1	11	7	3	0.5876	0.0468	0.2922	0.0280	0.0017

In Table A.3.1, in female rats, following the adjustment for multiplicity to get an overall rough 10% error rate and using the incidence in the no treatment group to decide if a tumor is rare or not, we would conclude, that the test of malignant sarcoma with bone formation was not quite statistically significant ($p = 0.0302 > 0.025$), though close. If one accepts the inflation of Type I error by the other tests, we would similarly conclude, that the test of trend in benign follicular cell adenoma of the thyroid in male rats was also not quite statistically significant ($p = 0.0385 > 0.025$). Again, using the incidence in the vehicle group to determine whether a tumor was rare or common, one would classify hepatocellular adenoma of the liver in female mice as a common tumor. Then, for this tumor the test of dose related trend over the first three groups was not statistically significant ($p = 0.0468 > 0.005$), as would be the pairwise comparison of the medium dose group to control ($p = 0.0280 > 0.01$). The only potentially statistically significant comparison is between the low dose and vehicle in hepatocellular adenoma of the liver in female mice ($p = 0.0017 < 0.01$) and that would come at the cost of inflation of the type I error to some level above 10%. In both rats and mice, no other test of trend or test of pairwise differences between the high dose and vehicle was statistically significant at a 0.05 level, let alone at one of the multiplicity adjusted significance levels.

Complete incidence tables are presented below. Note that for each organ tumor combination the first trend test is over all four study groups while the second trend test “trend/0-2” is over the first three groups, excluding the high dose group. The pairwise tests are the comparisons between actual dose groups and vehicle.

Table A.3.2 Incidence and Statistical Tests of Carcinogenicity in Male Rats

	Incidence				Significance Levels				
	Veh	Low	Med	Hi	Trend	0-2	High vs Veh	Medium vs Veh	Low vs Veh
Adipose, Brown									
M-Liposarcoma	0	1	0	0	0.7633	0.6818	.	.	0.5196
Adrenal, Medulla									
B&M Pheochromocytoma	11	8	10	14	0.2492	0.7301	0.4500	0.7702	0.8865
B-Pheochromocytoma	11	8	10	12	0.4093	0.7301	0.6048	0.7702	0.8865
M-Malignant Pheochromocytoma	0	0	0	2	0.0665	.	0.2724	.	.
Body, Whole/Cav									
M-Hemangiosarcoma	0	0	1	0	0.5096	0.3419	.	0.5196	.
M-Histiocytic Sarcoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
M-Lrg Granular Cell Leukemia	27	20	27	13	0.9950	0.6069	0.9990	0.6498	0.9577
M-Malignant Mesothelioma	1	2	1	2	0.4032	0.6381	0.5297	0.7671	0.5297
Gl, Mandib Saliv									
M-Carcinoma	0	0	0	1	0.2560	.	0.5196	.	.
M-Malignant Schwannoma	0	0	1	0	0.5096	0.3419	.	0.5196	.
Gl, Preputial									
B-Adenoma	0	0	1	0	0.5072	0.3377	.	0.5149	.
M-Carcinoma	3	1	1	0	0.9828	0.9237	1.0000	0.9458	0.9479
Jejunum									
M-Sarcoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Kidney									
B-Adenoma, Tubule Cell	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
M-Carcinoma, Tubule Cell	0	1	0	0	0.7633	0.6818	.	.	0.5196
Liver									
B-Adenoma, Hepatocellular	3	2	3	0	0.9537	0.6137	1.0000	0.6884	0.8364
B-Cholangioma	0	0	0	1	0.2560	.	0.5196	.	.
Mammary, Male									
M-Carcinoma	0	0	1	0	0.5096	0.3419	.	0.5196	.
Pancreas									
B-Adenoma, Islet Cell	3	1	0	2	0.7141	0.9900	0.8364	1.0000	0.9479
Parathyroid									
B-Adenoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Pituitary									
B-Adenoma	30	20	13	11	1.0000	0.9998	1.0000	0.9999	0.9902
Skin									
B-Adenoma, Sebaceous Gland	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
B-Fibroma	1	4	2	0	0.9043	0.4418	1.0000	0.5297	0.2067
B-Keratoacanthoma	1	2	1	1	0.6538	0.6439	0.7717	0.7717	0.5297
B-Neural crest tumor	0	0	0	1	0.2560	.	0.5196	.	.
B-Papilloma, Squamous Cell	0	0	0	2	0.0646	.	0.2675	.	.
M-Carcinoma, Squamous Cell	2	2	2	1	0.7690	0.6219	0.8892	0.7130	0.7130
M-Fibrosarcoma	2	4	0	1	0.8993	0.9062	0.8892	1.0000	0.3765
Testis									
B-Interstitial Cell Tumor	50	55	52	55	0.0678	0.2052	0.0810	0.2741	0.1537
Thyroid									
B-Adenoma, C-cell	12	7	8	7	0.8997	0.9042	0.9568	0.9182	0.9568
B-Adenoma, Follicular Cell	0	0	3	2	0.0879	0.0385	0.2675	0.1364	.
M-Carcinoma, C-cell	0	1	0	0	0.7633	0.6818	.	.	0.5196

Table A.3.3 Incidence and Statistical Tests of Carcinogenicity in Female Rats

	Incidence				Significance Levels				
	Veh	Low	Med	Hi	Trend	0-2	High vs Veh	Medium vs Veh	Low vs Veh
Adrenal, Medulla									
B-Pheochromocytoma	1	0	2	0	.	0.3226	1.0000	0.4370	1.0000
Body, Whole/Cav									
M-Carcinoma	0	1	0	0	0.7045	0.6414	.	.	0.4851
M-Histiocytic Sarcoma	0	2	0	1	0.3286	0.6291	0.3810	.	0.2378
M-Lrg Granular Cell Leukemia	13	12	11	7	0.6035	0.5296	0.6868	0.5748	0.5935
M-Malignant Mesothelioma	0	1	0	0	0.7045	0.6414	.	.	0.4851
Cervix									
B-Leiomyoma	0	0	1	0	0.4261	0.3034	.	0.4583	.
B-Polyp, Endometrial Stromal	0	3	1	0	0.6753	0.3287	.	0.4583	0.1106
M-Sarcoma, Endometrial Strom	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Gl, Clitoral									
B-Adenoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Gl, Zymbal's									
M-Carcinoma	0	1	0	0	0.7045	0.6414	.	.	0.4851
Heart									
B-Rhabdomyoma	0	0	0	1	0.1761	.	0.3735	.	.
Lung									
M-Carcinoma	0	1	0	0	0.7045	0.6414	.	.	0.4851
Mammary, Female									
B-Fibroadenoma	3	6	5	3	0.3471	0.2171	0.4150	0.2680	0.2147
M-Carcinoma	0	0	0	1	0.1808	.	0.3810	.	.
Nasal Turbinates									
M-Carcinoma, Squamous Cell	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Ovary									
B-Adenoma	0	0	0	1	0.1761	.	0.3735	.	.
Pancreas									
M-Carcinoma, Islet Cell	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Pituitary									
B-Adenoma	30	16	23	11	0.9243	0.7195	0.9849	0.7066	0.9944
M-Carcinoma	0	1	0	0	0.7045	0.6414	.	.	0.4851
Skin									
B-Fibroma	0	1	1	0	0.5256	0.3018	.	0.4639	0.4851
B-Lipoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
B-Neural crest tumor	0	0	1	0	0.4261	0.3034	.	0.4583	.
B-Papilloma, Squamous Cell	2	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
M-Carcinoma, Squamous Cell	0	1	0	0	0.7045	0.6414	.	.	0.4851
M-Fibrosarcoma	0	0	1	0	0.4261	0.3034	.	0.4583	.
M-Sarcoma	0	0	0	1	0.1761	.	0.3735	.	.
M-Sarcoma with bone formatio	0	0	0	2	0.0302	.	0.1366	.	.
Thyroid									
B-Adenoma, C-cell	8	3	4	4	0.6070	0.8906	0.7318	0.8935	0.9675
B-Adenoma, Follicular Cell	0	0	0	1	0.1761	.	0.3735	.	.
Urinary Bladder									
B-Papilloma, Transitional Ce	0	0	1	0	0.4261	0.3034	.	0.4583	.
M-Carcinoma, Transitional Ce	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Uterus									
B-Adenoma	0	3	0	1	0.3972	0.5849	0.3735	.	0.1106
B-Leiomyoma	0	0	1	1	0.1188	0.3034	0.3735	0.4583	.
B-Polyp, Endometrial Stromal	7	10	8	3	0.7181	0.3261	0.7939	0.3816	0.2392
M-Carcinoma	0	1	1	0	0.5235	0.2971	.	0.4583	0.4851
M-Sarcoma, Endometrial Strom	2	0	1	1	0.5398	0.8250	0.7595	0.8453	1.0000

Table A.3.4 Incidence and Statistical Tests of Carcinogenicity in Male Mice

	Incidence				Significance Levels					
	Veh	Low	Med	Hi	Trend	Trend High Medium Low			vs Veh vs Veh vs Veh	
						0-2	vs Veh	vs Veh		vs Veh
ADRENAL GLAND										
Benign phaeochromocytoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	
Subcapsular adenoma	12	9	10	12	0.4012	.	0.5400	0.7454	0.8093	
DUODENUM										
Adenoma	0	1	0	0	0.7433	0.6616	.	.	0.4924	
HARDERIAN GLAND										
Adenoma	2	3	2	2	0.5797	0.5909	0.6668	0.6904	0.4856	
KIDNEY										
Tubular adenoma	0	0	1	2	0.0576	.	0.2329	0.4962	.	
LIVER										
Cholangiocarcinoma	0	1	0	0	0.7433	.	.	.	0.4924	
Hepatocellular adenoma	7	11	13	10	0.2582	0.0891	0.2674	0.1055	0.2035	
Hepatocellular carcinoma	10	8	8	9	0.5203	0.7159	0.6239	0.7550	0.7441	
Ito cell tumour	0	0	1	0	0.4943	.	.	0.4962	.	
LUNG										
Adenocarcinoma	5	2	0	0	0.9995	0.9965	1.0000	1.0000	0.9376	
Adenoma	11	7	7	2	0.9945	0.8651	0.9983	0.8850	0.8783	
LYMPHORETICULAR SYSTEM										
Histiocytic sarcoma	3	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	
Malignant lymphoma (Lymphosarcoma)	5	3	2	1	0.9666	0.9139	0.9840	0.9401	0.8581	
Mast cell tumour	0	0	0	1	0.2443	.	0.4885	.	.	
PANCREAS										
Islet cell adenoma	0	1	0	1	0.3008	0.6616	0.4846	.	0.4924	
SPLEEN										
Haemangiosarcoma	0	0	2	0	0.4884	0.1100	.	0.2444	.	
SUBCUTANEOUS TISSUE										
Fibrosarcoma	1	0	1	0	0.8041	.	1.0000	0.7444	1.0000	
Systemic										
Haemangiosarcoma	1	0	3	1	0.3845	0.1844	0.7363	0.3096	1.0000	
Hemangioma/-sarcoma	1	0	3	1	0.3845	0.1844	0.7363	0.3096	1.0000	
TESTIS										
Benign Leydig cell tumour	1	0	0	0	1.0000	.	1.0000	1.0000	1.0000	
THYROID GLAND										
Follicular cell adenoma	0	1	0	0	0.7433	.	.	.	0.4924	
TONGUE										
Squamous cell carcinoma	2	0	0	0	1.0000	.	1.0000	1.0000	1.0000	
VASCULAR TISSUE										
Haemangiosarcoma	1	0	1	1	0.4400	.	0.7363	0.7519	1.0000	

Table A.3.4 Incidence and Statistical Tests of Carcinogenicity in Female Mice

	Incidence				Significance Levels					
	Veh	Low	Med	Hi	Trend	Trend High Medium Low			vs Veh vs Veh vs Veh	
						0-2	vs Veh	vs Veh		vs Veh
ADRENAL GLAND										
Benign phaeochromocytoma	2	1	0	1	0.7832	0.9606	0.8691	1.0000	0.8660	
Cortical adenocarcinoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	
BONE MARROW-FEMUR										
Haemangioma	0	0	0	1	0.2471	.	0.4884	.	.	
BRAIN										
Meningeal sarcoma	0	0	0	1	0.2471	.	0.4884	.	.	
CERVIX										
Squamous cell carcinoma	0	1	0	0	0.7422	0.6580	.	.	0.4884	
DUODENUM										
Adenoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	
HARDERIAN GLAND										
Adenoma	5	4	4	2	0.8812	0.6816	0.9376	0.7384	0.7302	

Table A.3.4 (cont.) Incidence and Statistical Tests of Carcinogenicity in Female Mice

	Incidence				Significance Levels				
	Veh	Low	Med	Hi	Trend	0-2	High vs Veh	Medium vs Veh	Low vs Veh
LIVER									
Hepatocellular adenoma	1	11	7	3	0.5876	0.0468	0.2922	0.0280	0.0017
Hepatocellular carcinoma	3	5	2	2	0.8093	0.7071	0.8126	0.8076	0.3331
Hepatocholeangiocellular carcinoma	0	0	1	0	0.4980	0.3333	.	0.4923	.
LUNG									
Adenocarcinoma	2	1	2	0	0.8978	0.5940	1.0000	0.6787	0.8660
Adenoma	5	2	4	0	0.9811	0.6971	1.0000	0.7384	0.9322
LYMPHORETICULAR SYSTEM									
Histiocytic sarcoma	3	5	5	1	0.8552	0.2778	0.9346	0.3336	0.3250
Malignant lymphoma (Lymphosarcoma)	23	23	18	17	0.9075	0.8311	0.8975	0.8574	0.5474
Mast cell tumour	0	1	0	0	0.7422	0.6580	.	.	0.4884
MAMMARY GLAND									
Adenocarcinoma	0	2	0	2	0.2027	0.6598	0.2404	.	0.2366
Adenoma	0	1	0	0	0.7412	0.6563	.	.	0.4844
ORAL CAVITY									
Squamous cell papilloma	0	0	0	1	0.2471	.	0.4884	.	.
OVARY AND OVIDUCT									
Benign granulosa/theca cell tumour	0	0	1	1	0.1848	0.3333	0.4884	0.4923	.
Cystadenocarcinoma	0	0	1	0	0.4980	0.3333	.	0.4923	.
Cystadenoma	0	0	1	1	0.1848	0.3333	0.4884	0.4923	.
Haemangioma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Tubulostromal adenoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
PANCREAS									
Islet cell adenoma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
PITUITARY GLAND									
Adenoma	18	17	24	18	0.4055	0.1229	0.5346	0.1448	0.5925
SKIN									
Sebaceous adenoma	0	2	1	0	0.7289	0.3623	.	0.4923	0.2327
Squamous cell papilloma	0	1	0	0	0.7412	0.6563	.	.	0.4844
SPINAL CORD									
Haemangiosarcoma	0	0	1	0	0.4980	0.3333	.	0.4923	.
SPLEEN									
Haemangiosarcoma	0	1	0	0	0.7412	0.6563	.	.	0.4844
STOMACH									
Squamous cell carcinoma	0	0	1	0	0.4980	0.3333	.	0.4923	.
Squamous cell papilloma	0	0	1	0	0.4980	0.3333	.	0.4923	.
Systemic									
Haemangioma	1	0	2	1	0.4136	0.3623	0.7402	0.4884	1.0000
Haemangiosarcoma	0	1	2	0	0.5863	0.1435	.	0.2404	0.4844
Hemangioma/-sarcoma	1	1	4	1	0.4521	0.1000	0.7402	0.1731	0.7361
TAIL									
Malignant schwannoma	0	1	0	0	0.7412	0.6563	.	.	0.4844
Squamous cell papilloma	0	0	1	0	0.4980	0.3333	.	0.4923	.
THYROID GLAND									
Follicular cell adenoma	0	0	0	1	0.2500	.	0.4923	.	.
UTERUS									
Fibroma	0	1	0	0	0.7412	0.6563	.	.	0.4844
Haemangioma	0	0	2	0	0.4961	0.1099	.	0.2404	.
Haemangiosarcoma	0	0	1	0	0.4980	0.3333	.	0.4923	.
Leiomyoma	0	0	1	0	0.4980	0.3333	.	0.4923	.
Leiomyosarcoma	0	1	0	2	0.1070	0.6563	0.2404	.	0.4844
Squamous polyp	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Stromal cell polyp	2	1	4	4	0.1373	0.2343	0.3255	0.3255	0.8660
Stromal cell sarcoma	1	0	1	1	0.4504	0.6598	0.7442	0.7442	1.0000
VAGINA									
Squamous cell papilloma	1	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
Stromal cell polyp	0	0	0	1	0.2500	.	0.4923	.	.

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.

De Iorio, M., Muller, P., Rosner, R.L., and MacEachern, S. (2004), "An ANOVA Model for Dependent Random Measures", Bayesian Nonparametric Nonproportional Hazards Survival Modeling", *Journal of the American Statistical Association*. **99**, 465, 205-215.

De Iorio, M., Johnson, W.O., Mueller, P., and Rosner, R.L. (2009), "Bayesian Nonparametric Nonproportional Hazards Survival Modeling", *Biometrics*, **65**, 3, 762-771.

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.

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.

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 (2001), *Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals* (DRAFT GUIDANCE), Center for Drug Evaluation and Research, Food and Drug Administration.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
IND-69416	ORIG-1	ASTELLAS PHARMA GLOBAL DEVELOPMENT INC	YM178

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
09/10/2010
Statistical Carcinogenicity Review

KARL K LIN
09/13/2010
Concur with review