

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*

**207947Orig1s000**

**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

## CARCINOGENICITY STUDY

**NDA:** 207947

**Drug Name:** UPTRAVI® (Selexipag)

**Indication:** Treatment of Pulmonary arterial hypertension.

**Sponsor:** Actelion Pharmaceuticals Ltd.  
c/o Actelion Clinical Research Inc.  
1820 Chapel Avenue West  
Cherry Hill, New Jersey 08002

**CRO:** [REDACTED] (b) (4)

**Date:** Data Submitted: 3 March 2015  
Assigned to Reviewer: 8 January 2015

**Review Priority:** Standard

**Biometrics Division:** Division 6

**Statistical Reviewer:** Steve Thomson

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

**Medical Division:** Division of Cardiovascular and Renal Products

**Toxicologist:** Reviewer: James M. Willard, Ph.D.  
Team Leader: Albert Defelice, Ph.D.

**Project Manager:** Wayne Amchin, RAC, MPA, M International Affairs

**Keywords:** Carcinogenicity, Cox regression, Kaplan-Meier product limit, survival analysis, Trend test

## Contents

<b>1. EXECUTIVE SUMMARY .....</b>	<b>3</b>
1.1. CONCLUSIONS AND RECOMMENDATIONS .....	3
1.2. BRIEF OVERVIEW OF THE STUDIES .....	7
1.3. STATISTICAL ISSUES AND FINDINGS .....	8
1.3.1. <i>Statistical Issues</i> .....	8
1.3.2. <i>Statistical Findings</i> .....	15
<b>2. INTRODUCTION .....</b>	<b>15</b>
2.1. OVERVIEW.....	15
2.2. DATA SOURCES .....	15
<b>3. STATISTICAL EVALUATION .....</b>	<b>15</b>
3.1. EVALUATION OF EFFICACY.....	15
3.2. EVALUATION OF SAFETY .....	15
3.2.1. STUDY NO. (b) (4) 5940: TWENTY-FOUR-MONTH ORAL GAVAGE CARCINOGENICITY STUDY OF NS-304 IN RATS.	15
3.2.2. STUDY NO. (b) (4) 5939: TWENTY-FOUR-MONTH ORAL GAVAGE CARCINOGENICITY STUDY OF NS-304 IN MICE.	22
<b>4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>29</b>
<b>5. SUMMARY AND CONCLUSIONS .....</b>	<b>29</b>
5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	29
5.2. CONCLUSIONS AND RECOMMENDATIONS .....	29
<b>APPENDICES: .....</b>	<b>30</b>
APPENDIX 1. SURVIVAL ANALYSES.....	30
APPENDIX 2. FDA POLY-K TUMORIGENICITY ANALYSIS .....	34
APPENDIX 3. REFERENCES .....	52

## 1. EXECUTIVE SUMMARY

Both the report for rats and the report for mice included the statement that: “The purpose of the present study is to evaluate carcinogenicity of NS-304 [(i.e., Uptravi®)] in rats [or mice] by a 2-year oral gavage administration. In addition, systemic exposure to NS-304 was evaluated.” (page 7 of both reports). Both studies were conducted at the (in English, rather unfortunately named) (b) (4)

### 1.1. Conclusions and Recommendations

The Sponsor summarizes study design in the rat report as follows: “Carcinogenicity of NS-304 (prostaglandin I<sub>2</sub> agonist) was evaluated in Crl:CD(SD) rats. NS-304 was administered orally by gavage to male and female rats (60/sex/group for a main group, 6 weeks of age at the start of treatment) for 104 weeks. The dosage levels of NS-304 were set at 0 (0.5 w/v% methylcellulose solution), 10, 30 and 100 mg/kg/day.” (page 14 of rat report) Counts of satellite study animals in both species are indicated by parentheses in Tables 1 and 2 below.!

Alternatively, gross aspects of the study design for the rat study are summarized in Table 1 below:

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

Treatment Group	Animals / Gender	Nominal Dose (mg/kg/day)	Concentration (mg/mL)
1. Vehicle <sup>1</sup>	60 ( 5)	0	0
2. Low	60 ( 8)	10	2
3. Medium	60 ( 8)	30	6
4. High	60 ( 8)	100	20

<sup>1</sup>0.5 w/v% methylcellulose solution (0.5 w/v% MC solution)

General aspects of the study design for the rather similar mice study are summarized in Table 2 below. The Sponsor’s report described the conduct of the mouse study as follows: “Carcinogenicity of NS-304 (prostaglandin I<sub>2</sub> agonist) was evaluated in B6C3F1/Crlj mice. NS-304 was administered orally by gavage to male and female mice (55/sex/group for a main group, 6 weeks of age at the start of treatment) for 2 years. The dosage levels of NS-304 were set at 0 (0.5 w/v% methylcellulose solution), 125, 250 and 500 mg/kg/day. Since the survival rate in females in the 500 mg/kg group decreased with the progression of the administration period, all surviving females in this group were prematurely sacrificed in Week 100 after administration for 99 weeks. Further, systemic exposure to NS-304 was evaluated by determining plasma NS-304 and its metabolite, MRE-269 concentrations on the satellite animals.” (page 14 of mice report)

**Table 2. Design of Mice Study** (dose volume 10 mL/kg)

Treatment Group <sup>1</sup>	# Main study animals (# toxicology study animals)/gender	Dose (mg/kg/day)	Concentration (mg/mL)
1. Vehicle <sup>1</sup>	55 (9)	0	0
2. Low	55 (24)	125	12.5
3. Medium	55 (24)	250	25
4. High	55 (29)	500	50

<sup>1</sup> 0.5% w/v% methylcellulose solution

Summary tables of survival in rats are presented in Tables 18 and 19, below, with corresponding tables in mice in tables 26 and 27. Kaplan-Meier survival curves for the both genders in both species are presented in Appendix 1, Figures A.1.1 through A.1.4. The interpretation of these plots is supported by the tests of homogeneity and differences in survival over the treatment groups. The statistical significance levels (i.e., p-values) for rats are provided in Table 3, below. One might note that the log rank tests place greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus, it actually tends to place more weight on differences in earlier events than does the log rank test.

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

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.5181	0.4880	0.1189	0.0656
No Trend over all four groups	0.3716	0.4084	0.0200	0.0095
No difference between high dose and vehicle	0.7931	0.8753	0.0444	0.0380

Figure A.1.1, in Appendix 1, the Kaplan-Meier estimated survival curves in male rats suggest that during the last third of the study the low dose group had slightly higher survival than the other three study groups, which in turn, were largely intertwined. However, none of the comparisons in male rats would be categorized as being statistically significant (i.e., all six  $p \geq 0.3716$ ). In almost a reversal of fate, in Figure A.1.2, in female rats, the the high dose group seems to have higher survival than the other study groups, but again with the remaining dose groups largely intertwined. For overall homogeneity, these differences are not sufficient to result in a statistically significant test of overall homogeneity, though close (Logrank  $p=0.1189$ , Wilcoxon  $p=0.0656$ ). However, there is evidence of a statistically significant test of trend in dose (Logrank  $p=0.0200$ , Wilcoxon  $p=0.0095$ ), and a somewhat weaker result in the test of differences between the high dose and vehicle control (Logrank  $p=0.0444$ , Wilcoxon  $p=0.0380$ ).

Similar results in mice are presented in Table 4 below.

**Table 4. Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice**

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.0002	0.0001	<0.0001	<0.0001
No Trend over all four groups	0.0100	0.0104	<0.0001	<0.0001
No difference between high dose and vehicle	0.0080	0.0083	<0.0001	<0.0001

In particular, from the Kaplan-Meier plot in Figure A.1.3., in male mice the vehicle control has, by a considerable extent, the lowest survival, with eventually the medium group having the highest survival, and the low and high dose groups eventually largely intertwined between these two curves. Note these differences are sufficient to result in statistically significant differences in tests for homogeneity in both genders (both  $p \leq 0.0002$ ), as is the test of trend in dose (both  $p \leq 0.0104$ ), as well as the test of no differences in the high and control low dose. Similarly, the comparison between the high dose and vehicle test of was statistically significant (both  $p \leq 0.0083$ ). In female mice the high dose group has much higher mortality than the other study groups, resulting in highly statistically significant tests of homogeneity, trend, and difference between high dose and control (all six  $p < 0.0001$ ). Note that even if the most conservative adjustment for multiplicity were applied, the results in female rats would still be highly statistically significant.

Typically a large number of tumors are identified in the analysis of neoplasms, implying a large number of statistical tests. Following the frequentist paradigm, when interpreting significance levels (i.e., p-values), one can use the Haseman-Lin-Rahman (HLR) rules to adjust for the multiplicity of tests. Two approaches have been investigated, one for testing dose related trend and pairwise comparison between the high dose and control separately and the other these hypotheses jointly (please see Section 1.3.1.5, below, for details). Usual statistical practice would be to test these hypotheses separately, but some scientists want to control Type I error only when simultaneously testing both the trend and pairwise hypotheses. That is, in the two year study, when testing for trend over dose and, separately, the difference between the highest dose group with a control group, to control the overall Type I error rate for the joint tests in a two species submission to roughly 10%, one compares the unadjusted significance level of the trend test to 0.005 for common tumors and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. For the testing these hypothesis jointly for common tumors one compares the unadjusted significance level of the trend test to 0.005 and the pairwise test to 0.05, and for rare tumors 0.025 for tests of trend and 0.10 the pairwise comparison. Using these adjustments for other tests, like testing the comparisons between the Low and Medium dose groups versus vehicle can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Tables 5 and 6 below show the tumors in rats and mice that had at least one non-multiplicity adjusted test that was statistically significant at or close, to a 0.10 level (or contributed to a significant test). For each tumor-organ combination the tumor incidence over the four dose groups is listed first, followed by the significance levels of the overall test of trend

over all four dose groups, and finally the comparison of the high, medium and low dose groups with vehicle.

**Table 5. Potentially Statistically Significant Results for Organ-Tumor Combinations in Rats**

Gender Organ/Tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Male Rats								
Testis								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.3				
LEYDIG CELL TUMOR	2	0	2	5	.0226	.2264	.7157	1
Thyroid								
# Evaluated	58	57	56	59				
Adj. # at Risk	44.9	47.4	46.4	45.6				
CARCINOMA, C-CELL	0	0	2	2	.0705	.2528	.2584	.
Female Rats								
Adrenal								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.5				
PHEOCHROMOCYTOMA	0	0	1	3	.0261	.1467	.4941	.
Adj. # at Risk	43.3	41.4	42.4	49.5				
Pheochromocytoma, Any	0	1	1	3	.0590	.1467	.4941	.4881
Thyroid								
# Evaluated	59	59	59	59				
Adj. # at Risk	42.5	40.9	42.9	49.5				
ADENOMA, C-CELL	1	5	5	7	.0999	.0478	.1008	.0899

Using the tumor incidence in the vehicle to determine whether a tumor should be classified as rare or common, only c-cell carcinoma and pheochromocytoma (both above) were classified as rare tumors, the remainder common. Although some of the p-values for these tumors fall below the 0.10 level, after adjusting for multiplicity using the Haseman-Lin-Rahman rules only the test of trend for Pheochromocytoma in female rats is close to statistical significance ( $p = 0.0261 \approx 0.025$ ). Complete tables of tumor incidence in both genders are given in Tables A.2.2 and A.2.3, in Appendix 2, below.

Similar results in mice are presented in Table 6, below.

**Table 6. Potentially Statistically Significant Results for Organ-Tumor Combinations in Mice**

organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Male Mice								
Testis								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
LEYDIG CELL TUMOR	0	0	1	2	.0644	.2816	.5484	.
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA, FOLLICULAR CELL	0	0	1	2	.0644	.2816	.5484	.
Adj. # at Risk	42.4	51.5	52.0	49.0				
Foll. Cell Adenoma/Carcinoma	0	0	2	2	.0749	.2816	.3034	.
Female Mice								
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.2	41.1				
ADENOMA, FOLLICULAR CELL	1	1	1	3	.0889	.2245	.7524	.7524

Only the tests of trend, above, show significance levels below 0.10 (i.e. 10%). Using the tumor incidence in the vehicle to determine whether a tumor should be classified as rare or common, only follicular cell adenoma in female mice would be classified as common, the remaining tumors as rare. But then, after adjusting for multiplicity using the Haseman-Lin-Rahman rules, no tests would be categorized as statistical significant.

Complete tables of tumor incidence are given in Tables A.2.5 and A.2.6, in Appendix 2, below.

### 1.2. Brief Overview of the Studies

Two studies were submitted, conducted at (b) (4) :

**Study No. (b) (4) 5940: Twenty-four-month oral gavage carcinogenicity study of NS-304 in rats.**

and,

**Study No. (b) (4) 5939: Twenty-four-month oral gavage carcinogenicity study of NS-304 in mice.**

These studies were designed to assess the carcinogenic potential of Uptravi (Selexipag). In both studies, the actual dose groups were labeled in this report as the Low, Medium, and High dose groups, respectively, plus the Vehicle control group.

### **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 comments on the details of 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 Wilcoxon statistic provided by SAS® (technically the Gehan-Wilcoxon statistic) can be cast as a log rank test weighted by the number of subjects at risk, and thus is more sensitive to earlier differences (when more subjects are at risk). The logrank test is most powerful when the survival curves track each other, and thus the hazards, i.e., the conditional probability of the event in the next infinitesimal interval, would be roughly proportional. Note the logrank test seems to be the test usually recommended by statisticians, and is one of the tests used by the Sponsor (in addition to the Tarone's test). Both the logrank and the Wilcoxon tests are used in the FDA analysis of mortality.

Appendix 1 reviews the specific FDA animal survival analyses in more detail. The results of the Sponsor's analysis are summarized in Sections 3.2.1.1 and 3.2.2.1, below..

##### **1.3.1.2. Multiplicity of Tests on Survival:**

Using both the logrank and Wilcoxon tests, for each gender in rats and, ignoring the positive control, mice there are six tests of survival differences in each gender in each species. Assuming tests were performed at the usual 0.05 level, and the tests were stochastically independent, but there were actually absolutely no differences in survival across groups (so one would hope no tests would be statistically significant), the probability of at least one statistically significant result in each gender in each species was about 0.2649 in each gender and 0.708 in both genders in both species. These bounds assume the tests are stochastically independent, which they clearly are not, but these values can give some idea of the possible price paid for the multiplicity of hypothesis tests in the statistical frequentist paradigm. Further general comments on adjusting for the multiplicity of tests are presented in Section 1.3.1.4 below.

##### **1.3.1.3. Tests on Neoplasms:**

Sponsors are requested to provide data in either SEND (Standard for the Exchange of Nonclinical Data) format, part of the CDISC consortium, or in the older FDA Biometrics format. Data from both studies fit the latter format. The FDA Biometrics format data sets requested for the analysis of rodent carcinogenicity studies are supposed to include a record for each animal organ combination that was not evaluated. If a number of the animals are not examined, but the proportions of animals showing the tumor under study in each treatment group is roughly the same as in the subset of animals actually reported the calculated p-values will generally be too large, i.e., results will be less statistically significant than they should be, possibly much less

The Sponsor's analyses of tumorigenicity in both species are Peto tests, with incidental and fatal plus mortality independent tumors. Note that Peto methods require accurate determination of whether a tumor is fatal or incidental. In both species survival was generally consistent across study dose groups.

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

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

#### **1.3.1.4. Multiplicity of Tests on Survival and Neoplasms:**

In each species and gender combination there were tests of homogeneity in survival or tumorigenicity over dose, tests of trend in survival, and comparisons of the high (and possibly other) dose to vehicle control. The individual p-values for hypothesis test are based on controlling the probability of rejecting a true null hypotheses in each separate test. There are a number of ways of controlling the overall error of rejecting any true null hypothesis among a set of such hypotheses. The most conservative test is based on so called Bonferroni comparisons where the individual p-value is divided by the number of comparisons. While experimenting with other approaches, current FDA practice in testing tumorigenicity is usually based on the Haseman-Lin-Rahman multiplicity adjustments..

The Haseman-Lin-Rahman rules are based on the original multiplicity adjustment of

Haseman (1983) and extended by Lin and Rahman on the basis of various simulations. Based on his extensive experience with such analyses, for pairwise tests in a two species study comparing control to the High dose group, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Lin & Rahman (1998) proposed a further p-value adjustment for tests of trend. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of trend, rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level. The general specifications are presented in the Table 4 below. 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).

The proposed Haseman-Lin-Rahman bounds are taken from *Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals*, (HHS, 2013). The bounds on the right in table 7, below, are grouped so that the last four columns correspond to testing both trend and the pairwise comparison between the high dose and control either separately or jointly. In this analysis we follow the usual practice of testing parameters separately, so the bounds in the leftmost columns are used. The observed tumor incidence in the vehicle group is used to decide if a tumor is classified rare or common.

**Table 7. Recommended Multiplicity Adjusted Bounds on Significance Levels**

Tests	Separate testing of trend and pairwise differences		Joint testing of trend and pairwise Differences	
	Trend	Pairwise	Trend	Pairwise
Common Tumor	0.005	0.01	0.005	0.05
Rare Tumor	0.025	0.05	0.025	0.10

In words, as noted in the FDA Guidance (2013) “For tests for positive trend alone, it is recommended that common and rare tumors are tested at 0.005 and 0.025 significance levels, respectively, in the two-year study ...

“For [the] control-high pairwise comparison alone, it is recommended that common and rare tumors are tested at 0.01 and 0.05 significance levels, respectively, in the two-year study ...

“For tests for positive trend and control-high pairwise comparison jointly, it is recommended that common and rare tumors are tested at 0.005 and 0.025 significance levels, respectively in trend test, and at 0.05 and 0.10 significance levels, respectively, in control-high pairwise comparison in the two-year study ...” (page 32 of 2013 Guidance)

The significance levels of the pairwise tests between the vehicle control with the Low and Medium dose groups are also provided in the tumor analysis tables below. Following the HLR

rules, adding these comparisons can be expected to increase the overall type I error rate to some level above the usual rough 10% level, possibly considerably larger. Again, because of the possibility of genetic drift and for convenience, incidence in the vehicle group is used to determine if the tumor is classified as rare or common.

#### 1.3.1.6. Validity of the Designs:

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

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

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

Lin and Ali (2006), quoting work by Haseman, have suggested that in standard laboratory rodent species, a survival rate of about 25 animals, out of 50 or more animals (i.e. 50%), between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. From tables 14 and 15 in Section 3.2.1.2 and tables 14 and 15 in Section 3.2.1.2 below, as a percentage of the High dose group animals that survived to week 91, this criterion is considerably exceeded in both genders (Male rats high dose: 71.7% and Female rats: 75.0%, Male mice high dose: 87.3% and Female mice: 61.8%). This may be interpreted as evidence that the MTD was not achieved in both genders in both species. However, such a determination requires the expertise of the toxicologist.

Tables 8 through 11 below indicate the weight changes in those animals that survived to near the end of each study. Chu, Ceuto, and Ward (1981), citing earlier work by Sontag et al. (1976) recommend that the MTD “is taken as ‘the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal’s natural life span’ ” From Tables 8 through 11 below, it seems that the weight criterion is exceeded in all dose groups in both genders in both species. This also may be further evidence that the MTD was exceeded in both species, but such a determination requires the expertise of the toxicologist.

**Table 8. Mean Weights and Changes (in g) in Male Rats**

Dose Group	Dose mg/kg/day	N	Day		Change from Baseline	% change relative to vehicle
			0	700		
1. Vehicle	0	34	236.0	842.4	606.5	
2. Low	10	41	238.2	771.8	533.5	88.0%
3. Medium	30	36	238.0	681.3	443.3	73.1%
4. High	100	31	236.43	627.75	391.3	64.5%

**Table 9. Mean Weights and Changes (in g) in Female Rats**

Dose Group	Dose mg/kg/day	N	Day		Change from Baseline	% change relative to vehicle
			0	700		
1. Vehicle	0	27	164.0	542.3	378.4	
2. Low	10	26	159.7	479.5	319.8	84.5%
3. Medium	30	27	161.7	427.0	265.3	70.1%
4. High	100	37	162.3	395.5	233.2	61.6%

The Sponsor's rat report notes the following:

“1) Male

In the 10 mg/kg group, statistically significant lower values for body weight were observed from Week 18 (Day 126) to the end of the administration period and mean body weight at the end of the administration period was 10% lower than that in the control group.

“In the 30 mg/kg group, lower values for body weight were observed from Week 1 (Day 7) to the end of the administration period and statistically significant differences were recorded at almost all measured points. Mean body weight at the end of the administration period was 21% lower than that in the control group.

In the 100 mg/kg group, statistically significant lower values for body weight were observed from Week 1 (Day 7) to the end of the administration period and mean body weight at the end of the administration period was 25% lower than that in the control group.

“2) Female

In the 10 mg/kg group, lower values for body weight were observed from Week 14 (Day 98) to the end of the administration period and statistically significant differences were recorded at many measured points. Mean body weight at the end of the administration period was 11% lower than that in the control group.

In the 30 mg/kg group, lower values for body weight were observed from Week 8 (Day 56) to the end of the administration period and statistically significant differences were recorded at almost all measured points. Mean body weight at the end of the administration period was 19% lower than that in the control group.

In the 100 mg/kg group, statistically significant lower values for body weight were observed in Week 1 (Day 7) and from Week 6 (Day 42) to the end of the administration period, and mean body weight at the end of the administration period was 27% lower than that in the control group.” (pages 39-40)

At least one way to indicate similar results in mice is as follows:

**Table 10. Mean Weights and Changes (in g) in Male Mice**

Dose Group	Dose mg/kg/day	N	Day		Change from Baseline	% change relative to vehicle
			0	700		
1. Vehicle	0	29	24.7	40.6	15.9	
2. Low	125	43	24.6	38.6	14.0	88.1%
3. Medium	250	49	24.7	36.3	11.6	73.0%
4. High	500	41	24.6	34.8	10.2	64.2%

**Table 11. Mean Weights and Changes (in g) in Female Mice**

Dose Group	Dose mg/kg/day	N	Day		Change from Baseline	% change relative to vehicle
			0	630		
1. Vehicle	0	50	19.6	37.4	17.8	
2. Low	125	51	19.6	34.1	14.5	81.5%
3. Medium	250	50	19.6	32.4	12.8	71.9%
4. High	500	34	19.5	31.4	11.9	66.9%

The Sponsor's mice report notes the following:

“1) Male

In the 125 mg/kg group, statistically significant lower values for body weight were observed from Week 10 (Day 70) to the end of the administration period and mean body weight at the end of the administration period was 5% lower than that in the control group.

In the 250 mg/kg group, statistically significant lower values for body weight were observed from Week 9 (Day 63) to the end of the administration period and mean body weight at the end of the administration period was 11% lower than that in the control group.

In the 500 mg/kg group, statistically significant lower values for body weight were observed in Week 1 (Day 7) and from Week 9 (Day 63) to the end of the administration period and mean body weight at the end of the administration period was 14% lower than that in the control group.

“2) Female

In the 125 mg/kg group, lower values for body weight were observed from around Week 26 (Day 182) to the end of the administration period and statistically significant differences were recorded at almost all measured points. Mean body weight at the end of the administration period was 8% lower than that in the control group. Although statistically significant differences were observed in Weeks 5 (Day 35) and 8 (Day 56), they were judged to be incidental, since they were temporary and minimal.

In the 250 mg/kg group, lower values for body weight were observed from around Week 4 (Day 28) to the end of the administration period and statistically significant differences were recorded at many measured points. Mean body weight at the end of the administration period was 11% lower than that in the control group.

In the 500 mg/kg group, lower values for body weight were observed from around Week 4 (Day 28) to the end of the administration period (Day 697) and statistically significant

differences were recorded at many measured points. Mean body weight at the end of the administration period was 14% lower than that in the control group.” (pages 39-40 of mouse report)

The Sponsor’s report also indicates that only the 10 mg/kg dose in rats was associated with food consumption comparable to the vehicle control, with lower consumption in the medium (30 mg/kg) and high (100 mg/kg) dose groups. In mice the higher dose groups were described as having lower food consumption during the first half or more of the study.

Again from 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. This suggests that a potentially useful way to assess whether or not the MTD was achieved is to measure early mortality not associated with any identified tumor. If this mortality is related to dose, it suggests that animals tend to die before having time to develop tumors. From the table below it seems that in rats there is no particular evidence of heterogeneity across dose groups.

**Table 12. Natural Death with No Identified Tumor in Rats (Male/Female)**

		1.Vehicle	2. Low	3.Medium	4.High
Males	Event	5	3	4	11
	No event	55	57	56	49
Females	Event	1	1	1	3
	No event	59	59	59	57

The apparent lack of homogeneity in natural death without tumor is confirmed the results of Fisher exact tests of homogeneity ( Males  $p = 0.0004$  and Females  $p = 0.0297$ ). This does seem to suggest that the high dose may have been above the MTD, however to be conclusive this observation requires the expertise of the toxicologist.

**Table 13. Natural Death with No Identified Tumor in Mice (Male/Female)**

		1.Vehicle	2. Low	3.Medium	4.High
Males	Event	12	4	4	6
	No event	43	51	51	49
Females	Event	1	4	3	32
	No event	54	51	52	23

In mice, there also seems to be an apparent lack of homogeneity in natural death without tumor in mice, although it is structurally different than the simple results in mice. In male mice the difference seems to be largely due to early deaths in the vehicle control, while in female mice has many more deaths without tumor in the high dose group. This is confirmed the results of Fisher exact tests of homogeneity ( Males  $p = 0.0773$  and Females  $p < 0.0001$ ). This does seem to suggest that at least in female mice the high dose may have been above the MTD, however, again, such a conclusion requires the expertise of the toxicologist

### 1.3.2. Statistical Findings

Please see Section 1.1 above.

## 2. INTRODUCTION

### 2.1. Overview

The Sponsor's reports summarize results from two two-year studies, one in rats, and the other in mice, both with daily gavage, to assess the carcinogenic potential of Uptravi (Selexipag) in the Sponsor's reports.

### 2.2. Data Sources

SAS data sets for both species, following the requested FDA Biostatistics format, both labeled tumor.sas7bdat, plus were translated from SAS transport files both labeled tumor.xpt. In addition both studies included SAS data sets labeled as food.sas7bdat and weights.sas7bdat as translated from the corresponding SAS transport files.

## 3. STATISTICAL EVALUATION

### 3.1. Evaluation of Efficacy

NA

### 3.2. Evaluation of Safety

More detailed results on the studies in rats and mice are presented below.

#### 3.2.1. Study No. <sup>(b)</sup><sub>(4)</sub> 5940: Twenty-four-month oral gavage carcinogenicity study of NS-304 in rats.

CRO: <sup>(b)</sup><sub>(4)</sub>

STUDY DURATION: 104 Weeks

DOSING STARTING DATE: December 13, 2007

STUDY COMPLETED (Terminal Necropsy): Males: December 11 and 14, 2009

Females: December 10, 2009. :

RAT STRAIN: Crl:CD(SD) Rats

ROUTE: Daily Oral gavage

The Sponsor's report describes the study as follows: "Two hundred and ninety-nine (299) male and 299 female Crl:CD(SD) SPF rats (<sup>(b)</sup><sub>(4)</sub>), at 5 weeks of age, were obtained on December 3, 2007. . . . The animals were quarantined and acclimatized for 10 days, and clinical observations (once a day) and measurement of body weight (3 times) were conducted. In addition, ophthalmological examination was conducted once on all animals. Based on the above examinations, healthy animals showing normal body weight gains and no abnormal clinical signs were selected and used at 6 weeks of age. During

the quarantine/acclimatization period, 1 female showed excoriation and was removed from animal allocation. Except this, no observable abnormalities were found in clinical observations or body weight; however, the animals with ophthalmological abnormalities were removed from the present study . . . .

“After animal selection, the animals were ranked by individual body weight on the day of animal allocation (2 days before the start of administration). Then the animals were assigned to each group so as to ensure the homogeneity of group means for body weight. This procedure was done using a computer by block placement and randomization methods (the appropriate number of groups was composed by block placement method, and test groups and animal numbers in each test group were randomly assigned).

“Two hundred and forty (240) animals of each sex were provided for a main group to evaluate toxicity and 29 animals of each sex were provided for a satellite group to evaluate toxicokinetics. Individual body weight on the starting day of administration ranged from 212 to 269 g in males and 142 to 184 g in females for the main group, and that for the satellite group ranged from 220 to 268 g in males and 145 to 184 g in females.

“Of the animals remained after animal allocation, 12 animals of each sex (total of 5 planned animals and 7 reserve animals) were reserved as monitor animals for microbiological test. Other animals were excluded from the study and they were used for collection of TK matrix.

“note: The number of animals ordered according to the protocol was 290 of each sex, but 299 of each sex” (pages 21-22 of rat report)

Gross aspects of the study designs for the main study animals are summarized in Table 14 below (a repeat of Table 1 above):

**Table 14. Design of Rat Study** (dose volume 5 mL/kg/day )

Treatment Group	Animals / Gender	Nominal Dose (mg/kg)	Concentration (mg/kg)
1. Vehicle <sup>1</sup>	60 ( 5)	0	0
2. Low	60 ( 8)	10	2
3. Medium	60 ( 8)	30	6
4. High	60 ( 8)	100	20

The Sponsor’s report indicates that dose levels were based upon results: “In a 26-week repeated dose toxicity study of NS-304 in rats with 4-week recovery period ( (b) (4) ), Study No. (b) (4) 5895, dosage levels: 0, 6, 25 and 100 mg/kg/day) . . . , suppression of body weight gain (male: -19%, female: -6%) and increased incidence or severity of diffuse acinar cell hyperplasia in the mammary gland was observed in females in the 100 mg/kg group. Therefore, the high dose level in the present study was set at 100 mg/kg/day and the middle and

low dose levels were set at 30 and 10 mg/kg/day, respectively. There were 4 test groups including a control group for each sex. Sixty (60) animals of each sex were provided for each test group as the main group. In addition, in each group, another 5 (control group) or 8 (dose groups) animals of each sex served as a satellite group for determination of plasma drug concentrations.” ( pages 23-24 of rat report)

Animals were housed individually with water available ad libitum. The Sponsor states that all animals were observed for clinical signs three times a day.

### 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 conclusions on survival in rats are summarized as follows:

“1) Male

There were no effects of NS-304 on the survival rate.

“2) Female

A statistically significant trend of an increase in the survival rate with increasing dose levels was observed, and the survival rate of the 100 mg/kg group was significantly higher than that of the control group.” (page 36 of rat report)!

**Table 15: Sponsor’s “Text Table 1. Summary of mortality and survival rate”**

Sex	Male				Female			
	0	10	30	100	0	10	30	100
Dose (mg/kg/day)								
No. of animals used	60	60	60	60	60	60	60	60
Week of administration								
1-26	0	1	0	1	0	0	0	0
1-52	5	3	2	3	2	0	2	1
1-78	11	7	8	10	10	16	13	4
1-105 <sup>a)</sup>	29	26	26	31	39	39	36	29
No. of survivors	31	34	34	29	21	21	24	31
Survival rate (%)	51.7	56.7	56.7	48.3	35.0\$	35.0	40.0	51.7*

Values in the table indicate the cumulative number of animals that died or were sacrificed as moribund.

\$: p<0.05 (statistically significant trend, Tarone.s test)

\*: p<0.05 (statistically significant difference from control group, log-rank test)

a): All surviving animals were necropsied in Week 105 after administration for 104 weeks.” (page 36 of rat report)

#### Tumorigenicity analysis:

The Sponsor’s report states that: “Tumors that occurred with high incidence (10 or more animals in total for each sex, >10) were evaluated using survival-adjusted Peto’s test (1) to assess increasing trend of incidence to dose level for all groups and to compare the incidence between the control group and each dose group. Tumors that occurred with low incidence (less than 10

animals in total for each sex, <10) were evaluated using exact permutation trend test to assess increasing trend of incidence to dose level for all groups and to compare the incidence between the control group and each dose group. For incidental tumors, the analysis intervals were: weeks 0 through 52, tumors and 0.025 (one tailed-level) for rare tumors. Pairwise comparison was conducted at the significance levels of 0.01 (one tailed-level) for common tumors and 0.05 (one tailed-level) for rare tumors. Common tumors were defined as those with a historical incidence in controls (CrI:CD(SD) rats) at (b) (4) exceeding 1% (>1%) and rare tumors as 1% or less (<1%),” (pages 34-35 of rat report)

The overall count of tumors is summarized in the following Table 16 (Text Table 11):

**Table 16: Sponsor’s “Text Table 11. Number of tumors and tumor bearers”**

Sex	Male				Female			
	0	10	30	100	0	10	30	100
Dose (mg/kg/day)								
No. of animals used	60	60	60	60	60	60	60	60
Total No. of tumors	121	124	91	84	159	153	128	125
No. of benign tumors	102	107	70	69	104	94	95	79
No. of malignant tumors	19	17	21	15	55	59	33	46
Total No. of tumor bearing animals	52	55	47	43	56	57	59	52
No. of benign tumor bearers	46	52	37	38	52	52	57	46
No. of malignant tumor bearers	18	17	18	13	29	28	23	27
No. of multiple tumor bearers	31	34	23	20	41	37	39	36

“There was no treatment-related increase in either number of tumors or tumor bearing animals in either sex.

“A decrease in the number of benign tumors was observed in males in the 30 and 100 mg/kg groups. Decreases in the total number of tumors, number of malignant tumor bearing animals and number of multiple tumor bearing animals were observed in males in the 100 mg/kg group. However, they were not regarded as an adverse effect of treatment, as these changes were decreases.” (page 45 of rat report)

“Tumor[s] that showed a increased incidence in dose groups as compared to the control group were observed in the testis and pituitary and incidences of testicular tumor and focal hyperplasia and pituitary tumor are summarized in the following . . . :”

**Table 17.Sponsor’s“Text Table 12. Incidence summary of major tumors and hyperplasias”**

Sex	Male				Female			
	0	10	30	100	0	10	30	100
Dose (mg/kg/day)	0	10	30	100	0	10	30	100
No. of animals used	60	60	60	60	60	60	60	60
Testis								
Leydig cell tumor	2\$ <sup>a</sup>	0	2	5	NA	NA	NA	NA
Hyperplasia, Leydig cell, focal (total, ±/+)	0	2	2	6	NA	NA	NA	NA
Pituitary								
Adenoma, anterior	20	40**	23	22	49	44	44	38

Numbers in the table indicate the number of animals with lesions.

\$: p<0.025 (statistically significant positive trend, rare tumor, Peto.s test)

\*\* : p<0.01 (statistically significant difference from the control group, common tumor, Peto.s test)

a): There was no statistically significant positive trend in the control, 10 and 30 mg/kg groups.

±: Minimal, +: Mild, NA: Not applicable” (page 46 of rat report)

The Sponsor’s report further summarizes results as:

“Testis:

“Marginally increased incidence of Leydig cell tumor was observed in the 100 mg/kg group, and statistically significant positive trend was noted (rare tumor, p<0.025); however, there was no statistical significance in pairwise comparison between the control and 100 mg/kg groups. Also, slightly higher incidence in focal hyperplasia of Leydig cells was observed in this group, although the difference in incidence between the two groups was very small. The tumor incidence of the 100 mg/kg group (5/60 animals, 8%) was marginally higher than that in our historical data (0 to 4% in the incidence). In addition, there was no statistically significant positive trend in incidence of Leydig cell tumor in the control, 10 and 30 mg/kg groups.

“In the pituitary, increased incidence of anterior adenoma was observed in males in the 10 mg/kg group with statistical significance (common tumor, p<0.01). However, it was not considered to be treatment related, since it was not dose-related.

“The other tumors . . . were judged to be incidental since they were consistent with the spectrum of spontaneous tumors expected in aged rats of this strain.” (pages 46-47 of rat report)

### 3.2.1.2 FDA Reviewer's Results

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

#### Survival analysis:

Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1, along with more details of the analysis. The following tables (Table 18 for male rats, Table 19 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 survived at the end of the interval.

**Table 18. Summary of Male Rats Mortality (dose/kg/day)**

Period	1.Vehicle	2.Low	3.Medium	4.High
0-50	5/60 91.7%	3/60 95.0%	2/60 96.7%	3/60 95.0%
51-70	2/55 88.3%	2/57 91.7%	2/58 93.3%	5/57 86.7%
71-90	10/53 71.7%	6/55 81.7%	13/56 71.7%	9/52 71.7%
91-105	12/43 51.7%	13/49 60.0%	9/43 56.7%	14/43 48.3%
terminal	31	36	34	29

<sup>1</sup> number deaths / number at risk

<sup>2</sup> per cent survival to end of period.

**Table 19. Summary of Female Rats Mortality (dose/kg/day)**

Period	1.Vehicle	2.Low	3.Medium	4.High
0-50	2/60 96.7%	0/60 100.0%	2/60 96.7%	1/60 98.3%
51-70	7/58 85.0%	8/60 86.7%	4/58 90.0%	2/59 95.0%
71-90	11/51 66.7%	16/52 60.0%	21/54 55.0%	12/57 75.0%
91-105	19/40 35.0%	15/36 35.0%	9/33 40.0%	14/45 51.7%
terminal	21	21	24	31

<sup>1</sup> number deaths / number at risk

<sup>2</sup> per cent survival to end of period.

Kaplan-Meier survival curves for the rat study are presented in Appendix 1. The results of statistical tests of differences in survival are given below (a repeat of Table 3):!

**Table 20. Statistical Significances of Tests of Homogeneity and Trend in Survival in Rats**

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.5181	0.4880	0.1189	0.0656
No Trend over all four groups	0.3716	0.4084	0.0200	0.0095
No difference between high dose and vehicle	0.7931	0.8753	0.0444	0.0380

From Figure A.1.1 in the appendix, in male rats the Kaplan-Meier estimated survival curves are largely intertwined, consistent with no tests of differences in survival being close to statistical significance. From Figure A.1.2 survival in female rats, the high dose group appears to have the highest survival with the other groups largely intertwined. For overall homogeneity the evidence for rejection is weak (Logrank  $p=0.1189$ , Wilcoxon  $p=0.0656$ ). There is evidence of a statistically significant test of trend in dose (Logrank  $p=0.0200$ , Wilcoxon  $p=0.0095$ ), and a somewhat weaker result in the test of differences between the high dose and vehicle control (Logrank  $p=0.0444$ , Wilcoxon  $p=0.0380$ ).

### Tumorigenicity analysis:

Table 21 below, a repeat of Table 6 above and Table A.2.1 below, shows the tumors in rats that had at least one non-multiplicity adjusted test that was statistically significant at a 0.10 or lesser level. For each tumor-organ combination the tumor incidence over the four dose groups is listed first, followed by the significance levels of the overall test of trend over all four dose groups, and finally the comparison of the high, medium and low dose groups with vehicle.

**Table 21. Potentially Statistically Significant Results for Organ-Tumor Combinations in Rats**

Gender Organ/Tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Male Rats								
Testis								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.3				
LEYDIG CELL TUMOR	2	0	2	5	.0226	.2264	.7157	1
Thyroid								
# Evaluated	58	57	56	59				
Adj. # at Risk	44.9	47.4	46.4	45.6				
CARCINOMA, C-CELL	0	0	2	2	.0705	.2528	.2584	.
Female Rats								
Adrenal								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.5				
PHEOCHROMOCYTOMA	0	0	1	3	.0261	.1467	.4941	.
Adj. # at Risk	43.3	41.4	42.4	49.5				
Pheochromocytoma, Any	0	1	1	3	.0590	.1467	.4941	.4881
Thyroid								
# Evaluated	59	59	59	59				
Adj. # at Risk	42.5	40.9	42.9	49.5				
ADENOMA, C-CELL	1	5	5	7	.0999	.0478	.1008	.0899

Using the tumor incidence in the vehicle to determine whether a tumor should be classified as rare or common, only c-cell carcinoma and pheochromocytoma (both above) were classified as rare tumors, the remainder common. Although some of these tumors exceed the 0.10 level, after adjusting for multiplicity using the Haseman-Lin-Rahman rules only the test of trend for Pheochromocytoma is close to statistical significance ( $p = 0.0261 \approx 0.025$ ).

Complete tables of tumor incidence are given in Tables A.2.2 and A.2.3, in Appendix 2, below.

### 3.2.2 Study No. (b) (4) 5939: Twenty-four-month oral gavage carcinogenicity study of NS-304 in mice.

CRO: (b) (4)

STUDY DURATION: Up to 104 Weeks

DOSING STARTING DATE: December 28, 2007

STUDY COMPLETED: February 9, 2011

MOUSE STRAIN: SB6C3F1/Crlj SPF mice

ROUTE: Daily Oral gavage

The Sponsor's report indicates that in the mouse study: "Carcinogenicity of NS-304 (prostaglandin I2 agonist) was evaluated in B6C3F1/Crlj mice. NS-304 was administered orally by gavage to male and female mice (55/sex/group for a main group, 6 weeks of age at the start of treatment) for 2 years. The dosage levels of NS-304 were set at 0 (0.5 w/v% methylcellulose solution), 125, 250 and 500 mg/kg/day. Since the survival rate in females in the 500 mg/kg group decreased with the progression of the administration period, all surviving females in this group were prematurely sacrificed in Week 100 after administration for 99 weeks. Further, systemic exposure to NS-304 was evaluated by determining plasma NS-304 and its metabolite, MRE-269 concentrations on the satellite animals." (page 14 of mouse report)

General aspects of the study design for the mice study are also summarized in Table 18 below (a repeat of Table 2 above):

**Table 22. Design of Mice Study (Volume 10 mL/kg)**

Treatment Group <sup>1</sup>	# Main study animals (# toxicology study animals)/gender	Dose (mg/kg/day)	Concentration (mg/mL)
1. Vehicle <sup>1</sup>	55 (9)	0	0
2. Low	55 (24)	125	12.5
3. Medium	55 (24)	250	25
4. High	55 (29)	500	50

<sup>1</sup>0.5% w/v% methylcellulose solution

The Sponsor's report indicates that the oral route was selected since it was an intended route in clinical use. The length of the administration period was 24 months (104 weeks, until the day before necropsy) according to the toxicity guideline. In females in the 500 mg/kg group, since the mortality increased with the progression of the administration period, all surviving females in this group were prematurely sacrificed in Week 100 after administration for 99 weeks when survivors decreased to 15 (73% mortality). The frequency of administration was once

daily (7 times per week), which is ordinary for repeated dose toxicity studies. The dose volume was set at 10 mL/kg body weight and dosing formulations were administered orally once daily (between 08:38 and 15:10) by gavage using a flexible stomach tube. Animals in the control group received the vehicle (0.5 w/v% MC solution) in the same manner. Individual dose volumes (unit: 0.01 mL) were calculated based on the most recently measured body weight.” (page 24 of mouse report)

Animals were approximately six weeks old at first dosing. Animals were housed individually with food and water available ad libitum. The Sponsor states that animals were checked three times daily.

### **3.2.2.1 Sponsor’s Results and Conclusions**

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

#### **Survival analysis:**

The Sponsor’s report notes that “The survival rate in females in the 500 mg/kg group was extremely low, where many deaths occurred in the late phase of the administration period. Contrary to females, in males, higher survival rate was observed in all dose groups.” (page 14 of mouse report)

Specifically, mortality and survival rate are summarized in the following table:

**Table 23: Text Table 1. Summary of mortality and survival rate**

Sex	Male				Female			
	0	125	250	500	0	125	250	500 <sup>a)</sup>
Dose (mg/kg/day)								
No. of animals used	55	55	55	55	55	55	55	55
Week of administration								
1-26	0	0	0	3	0	1	0	3
1-52	2	0	1	3	0	1	0	3
1-78	11	2	4	5	1	1	2	9
1-100	26	12	6	14	8	5	12	40
1-105 <sup>b)</sup>	28	16	9	15	13	9	16	NA
No. of survivors	27	39	46	40	42	46	39	15
Survival rate (%)								
Week 100	52.7	78.2	89.1	74.5	85.5	90.9	78.2	27.3*
Week 105b)	49.1\$	70.9*	83.6*	72.7*	76.4\$	83.6	70.9	NA

a): All surviving females in the 500 mg/kg group were sacrificed in Week 100 after administration for 99 weeks.

b): All surviving animals except for females in the 500 mg/kg group were necropsied as scheduled in Week 105 after administration for 104 weeks.

Values in the table indicate cumulative number of animals that died or were sacrificed as moribund.

\$:  $p < 0.05$  (statistically significant trend, Tarone's test)

\*:  $p < 0.05$  (statistically significant difference from the control group, log-rank test)

NA: Not applicable

#### “1) Male

A statistically significant trend of an increase in the survival rate with increasing dose levels was observed, and the survival rates in all dose groups were significantly higher than that in the control group.

#### “2) Female

A statistically significant trend of a decrease in the survival rate with increasing dose levels was observed, and the survival rate in the 500 mg/kg group was extremely low with statistical significance, where many deaths occurred in the late phase of the administration period.” (page 36 of mouse report)

### **Tumorigenicity analysis:**

The Sponsor's analysis of carcinogenicity is based on Peto survival-adjusted tests. Where results with low tumor incidence were computed using exact permutation tests. In particular The Sponsor claims that the (b) (4)

**Table 24: Text Table 9. Number of tumors and tumor bearers**

Sex	Male				Female			
	0	125	250	500	0	125	250	500 <sup>a)</sup>
Dose (mg/kg/day)	0	125	250	500	0	125	250	500 <sup>a)</sup>
No. of animals used	55	55	55	55	55	55	55	55
Total No. of tumors	62	46	44	35	89	81	66	30
No. of benign tumors	37	24	23	21	45	40	27	21
No. of malignant tumors	25	22	21	14	44	41	39	9
Total No. of tumor bearing animals	38	31	26	27	45	41	42	19
No. of benign tumor bearers	24	16	16	16	31	26	23	15
No. of malignant tumor bearers	21	18	17	13	32	30	32	8
No. of multiple tumor bearers	16	13	10	6	6	23	19	9

<sup>a)</sup> All surviving females in the 500 mg/kg group were sacrificed prematurely in Week 100 after administration for 99 weeks.

“There was no treatment-related increase in either number of tumors or tumor bearing animals in either sex.

“Decreases in total number of tumors and number of benign tumors were observed in females in the 250 and 500 mg/kg groups. Decreases in number of malignant tumors, total number of tumor bearing animals, number of benign tumor bearing animals and number of malignant tumor bearing animals were observed in females in the 500 mg/kg group. A decrease in number of multiple tumor bearing animals was observed in both sexes in the 500 mg/kg group. However, they were not regarded as an adverse effect of treatment, as these changes were decreases.” (page 45 of mouse report)

“Tumors that showed a tendency toward increase in the incidence in the dose groups as compared to the control group were observed in the thyroid and the incidence of tumors and related non-tumor findings are summarized in the following [table]:”

**Table 25: Text Table 10. Incidence summary of tumors and hyperplasia in the thyroid**

Sex	Male				Female			
	0	125	250	500	0	125	250	500 <sup>a)</sup>
Dose (mg/kg/day)								
No. of animals used	55	55	55	55	55	55	55	55
Thyroid								
Adenoma, follicular cell	0	0	1	2	1	1	1	3
Carcinoma, follicular cell	0	0	1	0	0	0	0	0
Adenoma + Carcinoma, follicular cell	0	0	2	2	1	1	1	3
Hyperplasia/hypertrophy, follicular cell (total)	1	2	17	36	2	51	52	52
(±)	1	2	17	33	2	41	20	4
(+ / ++)	0	0	0	3	0	10	32	48

<sup>a)</sup>: All surviving females in the 500 mg/kg group were sacrificed prematurely in Week 100 after administration for 99 weeks.

Numbers in the table indicate the number of animals with lesions.

±; Minimal, +; Mild, ++; Moderate

“Thyroid:

“Slightly higher incidence of follicular cell tumors (adenoma + carcinoma) were observed in 2 males each in the 250 and 500 mg/kg groups and in 3 females in the 500 mg/kg group, although there were no statistically significant differences in either trend analysis or pairwise comparison between the control and any dose group. At the same time, increased incidence and severity of hyperplasia/hypertrophy of the follicular cells were observed in the above groups. Follicular cell adenoma was also observed in 1 female each in the 125 and 250 mg/kg groups, respectively, and the possibility of involvement of treatment could not be completely excluded, since increased incidence and severity of hyperplasia/hypertrophy of the follicular cells were also observed in these groups. However, it was hardly judged to be treatment-related since follicular cell adenoma was also observed in 1 control female in the present study.

“Decreased incidences of tumors were observed in the following; however, reduced incidences of common spontaneous tumors are not regarded as adverse effects of treatment.

Liver: Hepatocellular adenoma in both sexes in the 250 and 500 mg/kg groups

Pituitary: Anterior adenoma in females in the 500 mg/kg group

Hemolymphoreticular: Malignant lymphoma in females in the 500 mg/kg group

“The other tumors described in the Tables and Appendices were judged to be incidental since they were consistent with the spectrum of spontaneous tumors expected in aged mice of this strain.” (page 46 of mouse report)

**3.2.1.2 FDA Reviewer's Results**

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

**Survival analysis:**

Kaplan-Meier plots comparing treatment groups in are given in Appendix 1, along with more details of the analysis. The following tables (Table 26 for male mice and Table 27 for female mice) summarize the mortality results for the dose groups. As with rats, 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 specified interval.

**Table 26. Summary of Male Mice Mortality**

Period	1.Vehicle	2.Low	3.Medium	4.High
0-50	1/55 98.2%	0/55 100.0%	1/55 98.2%	3/55 94.5%
51-70	9/54 81.8%	1/55 98.2%	2/54 94.5%	2/52 90.9%
71-90	9/45 65.5%	5/54 89.1%	1/52 92.7%	2/50 87.3%
91-105	9/36 49.1%	10/49 70.9%	5/51 83.6%	8/48 72.7%
terminal	27	39	46	40

<sup>1</sup> number deaths / number at risk

<sup>2</sup> per cent survival to end of period.

**Table 27. Summary of Female Mice Mortality**

Period	1.Vehicle	2.Low	3.Medium	4.High
0-50	0/55 100.0%	1/55 98.2%	0/55 100.0%	3/55 94.5%
51-70	1/55 98.2%	0/54 98.2%	0/55 100.0%	3/52 89.1%
71-90	4/54 90.9%	3/54 92.7%	5/55 90.9%	15/49 61.8%
91-105	8/50 76.4%	5/51 83.6%	11/50 70.9%	19/34 27.3%
terminal	42	46	39	15 <sup>1</sup>

<sup>1</sup> number deaths / number at risk

<sup>2</sup> per cent survival to end of period.

**Table 28. Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice**

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.0002	0.0001	<0.0001	<0.0001
No Trend over all four groups	0.0100	0.0104	<0.0001	<0.0001
No difference between high dose and vehicle	0.0080	0.0083	<0.0001	<0.0001

In Appendix 1, from Figure A.1.3 in male mice the vehicle control has, by a considerable extent, the lowest survival, with eventually the medium group having the highest survival, and the low and high dose groups eventually largely intertwined between these two curves. Note these differences are sufficient to result in statistically significant differences in tests for homogeneity in both genders (both  $p \leq 0.0002$ ), as is the test of trend in dose (both  $p \leq 0.0104$ ), as well as the test of no differences in the high and control low doses. Similarly, the comparison between the high dose and vehicle test of was statistically significant (both  $p \leq 0.0083$ ). In female mice the high dose groups has higher mortality than the other study groups, resulting in highly statistically significant tests of homogeneity, trend, and difference between high dose and control (all six  $p < 0.0001$ ).

Again, Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1, along with more details of the survival analysis.

#### **Tumorigenicity analysis:**

Table 29 below, a repeat of Table 7 above and Table A.2.2 below, shows the organ-tumor combinations associated with at least one non-multiplicity adjusted test that was statistically significant at a 0.10 level.

**Table 29. Potentially Statistically Significant Results for Organ-Tumor Combinations in Mice**

organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Male Mice								
Testis								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
LEYDIG CELL TUMOR	0	0	1	2	.0644	.2816	.5484	.
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA,FOLLICULAR CELL	0	0	1	2	.0644	.2816	.5484	.
Adj. # at Risk	42.4	51.5	52.0	49.0				
Foll. Cell Adenoma/Carcinoma	0	0	2	2	.0749	.2816	.3034	.
Female Mice								
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.2	41.1				
ADENOMA,FOLLICULAR CELL	1	1	1	3	.0889	.2245	.7524	.7524

Using the tumor incidence in the vehicle to determine whether a tumor should be classified as rare or common, only follicular cell adenoma would be classified as common, the remainder rare. Again, several of the tests of trend fall below the 0.10 level, after adjusting for multiplicity using the Haseman-Lin-Rahman rules no tests would be categorized as statistical significant. Complete tables of tumor incidence are given in Tables A.2.5 and A.2.6, in Appendix 2, below.

#### 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

#### 5. SUMMARY AND CONCLUSIONS

##### 5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

##### 5.2. Conclusions and Recommendations

Please see section 1.1 above.

**APPENDICES:****Appendix 1. Survival Analyses**

Simple summary life tables in mortality in rats are presented in the report (Tables 18 and 19, above). Kaplan-Meier estimated survival curves across study groups for each gender in rats are displayed below in Figures A.1.1 and A.1.2. The 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 treatment groups. The statistical significance levels (i.e., p-values) are provided in Table A.1.1., below. One might note that the log rank tests place greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus, it actually tends to place more weight on differences in earlier events than does the log rank test.

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

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.5181	0.4880	0.1189	0.0656
No Trend over all four groups	0.3716	0.4084	0.0200	0.0095
No difference between high dose and vehicle	0.7931	0.8753	0.0444	0.0380

Kaplan-Meier survival curves for these studies are presented below. From Figure A.1.1, the Kaplan-Meier estimated survival curves in male rats suggest that during the last third of the study the low dose group had slightly higher survival than the other three study group, which in turn, were largely intertwined. However, none of the comparisons in male rats would be categorized as being statistically significant (i.e., all six  $p \geq 0.3716$ ). In almost a reversal of fate, in Figure A.1.2, in female rats, the the high dose group seems to have higher survival than the other study groups, but again with the remaining dose groups largely intertwined. These differences are not sufficient to result in a statistically significant, though close to significance, test of overall homogeneity (Logrank  $p=0.1189$ , Wilcoxon  $p=0.0656$ ). However, there is evidence of a statistically significant test of trend in dose (Logrank  $p=0.0200$ , Wilcoxon  $p=0.0095$ ), and a somewhat weaker result in the test of differences between the high dose and vehicle control (Logrank  $p=0.0444$ , Wilcoxon  $p=0.0380$ ).

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

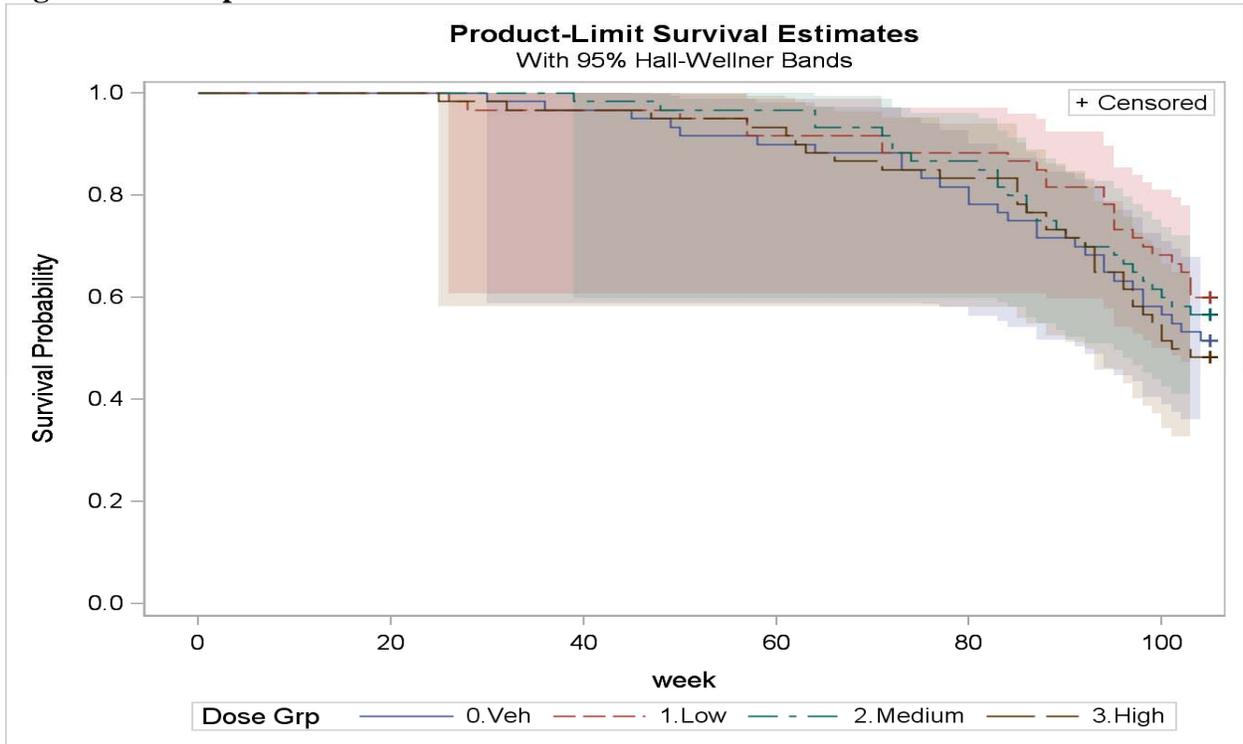
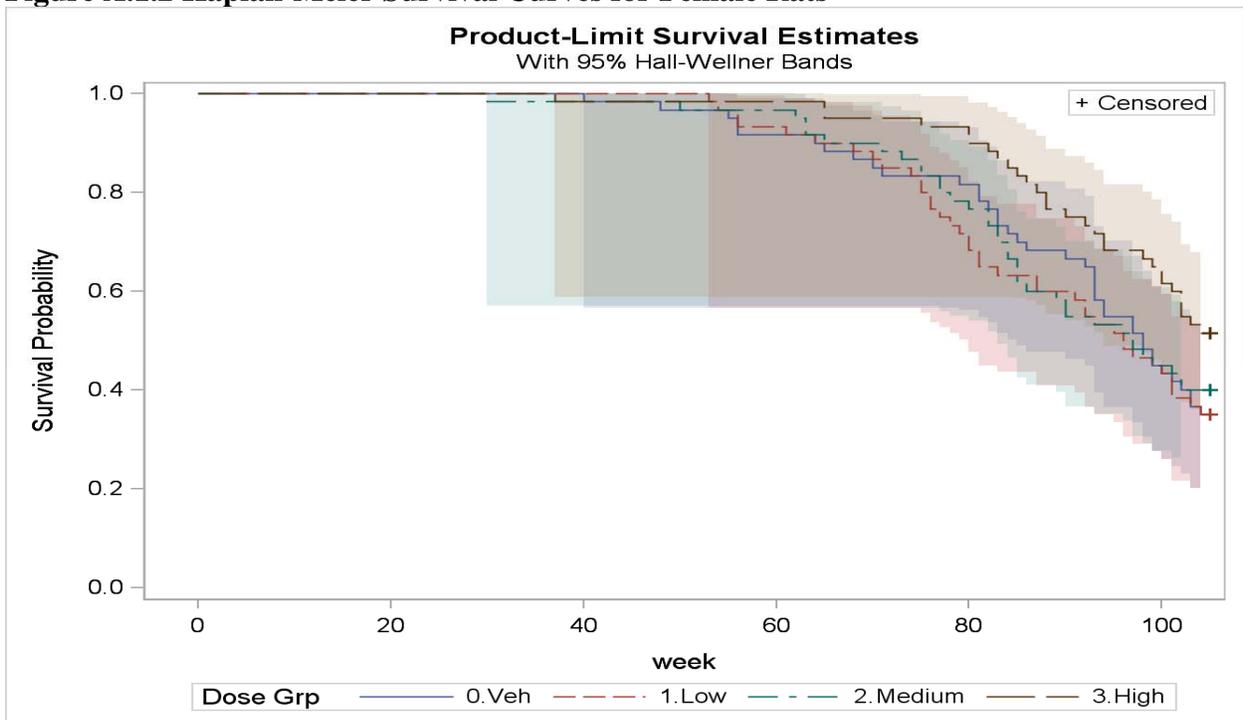


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



The results of the same set of statistical comparisons in survival are given in Table A.1.2., with corresponding Kaplan-Meier survival curves in Figures A.1.3 and A.1.4, below. In mice a there is a somewhat different pattern. Note that summary tables of survival are given in Tables 26 and 27 above. .

**Table A.1.2. Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice**

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.0002	0.0001	<0.0001	<0.0001
No Trend over all four groups	0.0100	0.0104	<0.0001	<0.0001
No difference between high dose and vehicle	0.0080	0.0083	<0.0001	<0.0001

From Figure A.1.3 in male mice the vehicle control has, by a considerable extent, the lowest survival, with eventually the medium group having the highest survival, and the low and high dose groups eventually largely intertwined between these two curves. Note these differences are sufficient to result in statistically significant differences in tests for homogeneity in both the logrank and Wilcoxon tests (both  $p \leq 0.0002$ ), as is the test of trend in dose (both  $p \leq 0.0104$ ), as wells as the test of no differences in the high and control low doses. Similarly, the comparison between the high dose and vehicle test of was statistically significant (both  $p \leq 0.0083$ ). In female mice the high dose groups have much higher mortality than the other study groups, sufficient to result in highly statistically significant tests of homogeneity, trend, and difference between high dose and control (all six  $p < 0.0001$ ).

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

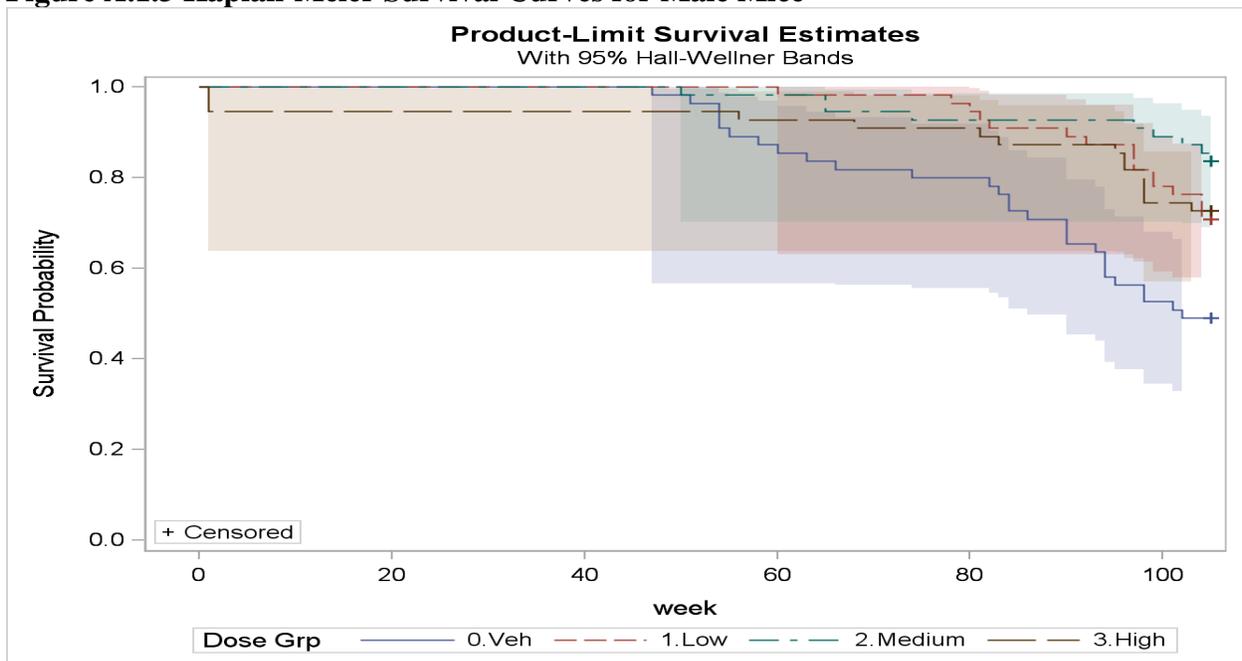
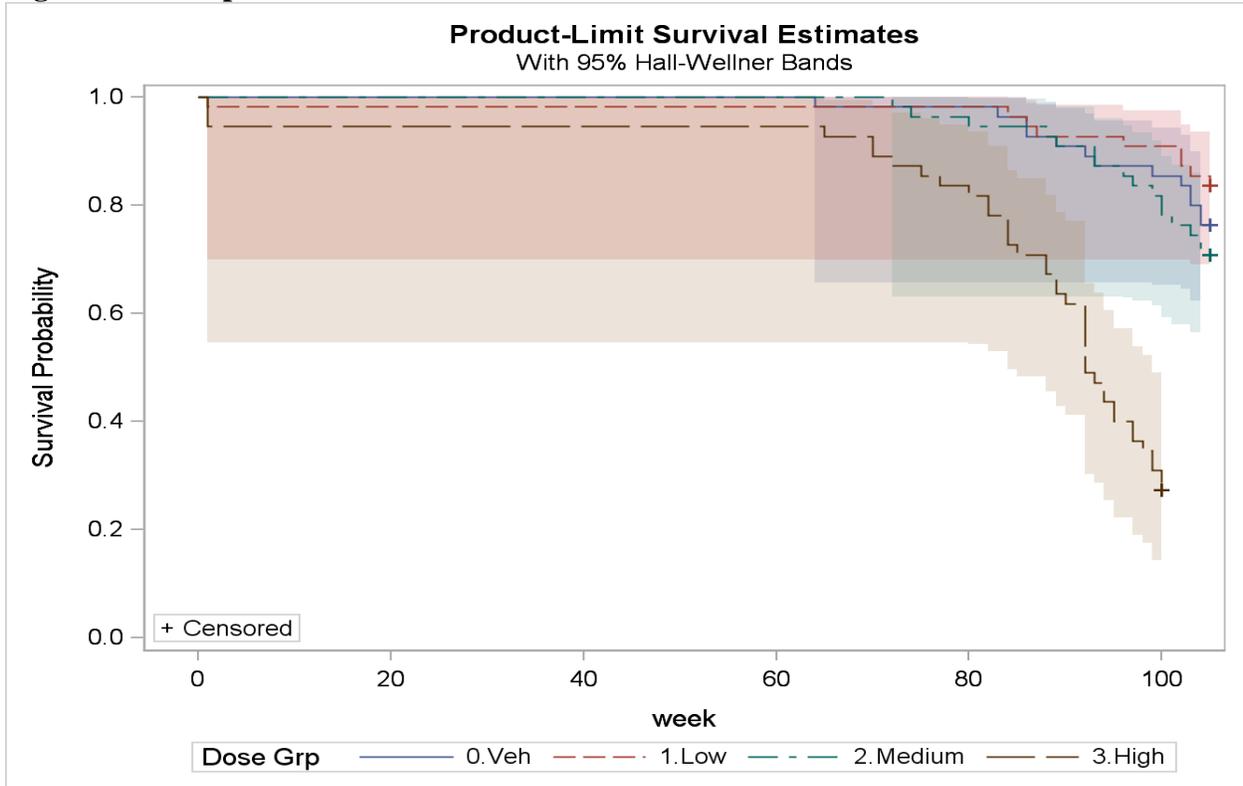


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



## Appendix 2. FDA Poly-k Tumorigenicity Analysis

The poly-k test, here with  $k = 3$ , modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The tests used here are small sample exact permutation tests of tumor incidence. When there were no tumors of the specific type being analyzed in either column of the 2x2 table corresponding to a pairwise comparison an argument could be made that the p-value for this test should be 1.0. However, largely for readability, in the tables below these p-values are considered as missing (i.e., corresponding to a null test), denoted by a period “.”. Note that the StatXact program used for these analyses adjusts for the variance, which would be 0. Then the significance levels of the test statistics are based on the result of a division by 0, i.e., undefined, and hence StatXact codes these p-values as missing.

For each gender by organ combination the number of animals microscopically analyzed is presented first. Note that indicating an organ was not examined requires a specification in the data (please see section 2.2 above). It is possible that this specification could be missing for some organ combinations in some study groups in this data. Then the number of animals at risk could be inflated, and the proportion of animals with tumor would be artificially decreased. Thus, as discussed in Section 1.3.1.5 above, for some of these organs it is possibly more appropriate to define the actual endpoint used in the statistical analysis be the condition of being microscopically analyzed and show the tumor. This does have problems if the treatment groups are treated equally except for actual treatment applied.

The entry for each tumor is preceded by the adjusted number of animals at risk for that endpoint. It seems clear that an animal that dies early without having displaying that endpoint reduces the size of the risk set for that getting that particular endpoint. The poly-k test down weights such animals, and as also discussed in Section 1.3.1.5, above, the sum of these poly-k weights seems to be a better estimate of the number of animals at risk of getting that tumor than the simple number of animals analyzed. This sum is given in the row labeled “Adjusted # at risk”. For these analyses, incidence in the control vehicle, water group is used to assess background tumor incidence, and thus whether a tumor is considered to be rare (background incidence < 1%) or common. Note that for these analyses a tumor is only classified as rare if the vehicle control group shows none of that particular tumor.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman (HLR) rules discussed in Section 1.3.1.6 are often applied. In this particular case we have two two-year study in rats and mice. An adjustment that seems to work is that for a roughly 10% overall error rate tests of trend would be considered significant if the tests for positive trend alone would be tested at 0.005 and 0.025 significance levels, for common and rare tumors respectively. Control-high pairwise differences would be tested at a 0.01 and 0.05 significance levels, for common and rare tumors respectively. If we require both the tests of trend and the pairwise comparison to be significant, the only change would be that the pairwise test in the two year study be tested at a 0.10 level for rare tumors. Using these adjustments for other tests, like testing the comparisons

between the low and medium dose groups versus vehicle can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Table A.2.1, below, a repeat of Table 5 above, shows those rows with at least one tumor with at least one non-multiplicity adjusted test that was statistically significant or close, to a 0.10 level.

**Table A.2.1. Potentially Statistically Significant Results for Organ-Tumor Combinations in Rats**

Gender Organ/Tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	phigh	pmed	plow
					vsVeh	vsVeh	vsVeh	
Male Rats								
Testis								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.3				
LEYDIG CELL TUMOR	2	0	2	5	.0226	.2264	.7157	1
Thyroid								
# Evaluated	58	57	56	59				
Adj. # at Risk	44.9	47.4	46.4	45.6				
CARCINOMA, C-CELL	0	0	2	2	.0705	.2528	.2584	.
Female Rats								
Adrenal								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.5				
PHEOCHROMOCYTOMA	0	0	1	3	.0261	.1467	.4941	.
Adj. # at Risk	43.3	41.4	42.4	49.5				
Pheochromocytoma, Any	0	1	1	3	.0590	.1467	.4941	.4881
Thyroid								
# Evaluated	59	59	59	59				
Adj. # at Risk	42.5	40.9	42.9	49.5				
ADENOMA, C-CELL	1	5	5	7	.0999	.0478	.1008	.0899

Using the tumor incidence in the vehicle to determine whether a tumor should be classified as rare or common, only c-cell carcinoma and pheochromocytoma (both above) were classified as rare tumors, the remainder common. Although some of these tumors exceed the 0.10 level, after adjusting for multiplicity using the Haseman-Lin-Rahman rules only the test of trend for Pheochromocytoma is close to statistical significance ( $p = 0.0261 \approx 0.025$ ). Complete tables of tumor incidence are given in Tables A.2.3 and A.2.4, below.

Similar results in mice are presented in Table A.2.2, below (also a repeat of Table 6 above):

**Table A.2.2. Potentially Statistically Significant Results for Organ-Tumor Combinations in Mice**

organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Male Mice								
Testis								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
LEYDIG CELL TUMOR	0	0	1	2	.0644	.2816	.5484	.
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA,FOLLICULAR CELL	0	0	1	2	.0644	.2816	.5484	.
Adj. # at Risk	42.4	51.5	52.0	49.0				
Foll. Cell Adenoma/Carcinoma	0	0	2	2	.0749	.2816	.3034	.
Female Mice								
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.2	41.1				
ADENOMA,FOLLICULAR CELL	1	1	1	3	.0889	.2245	.7524	.7524

Using the tumor incidence in the vehicle to determine whether a tumor should be classified as rare or common, only follicular cell adenoma would be classified as common, the remainder rare. Again, several of the tests of trend fall below the 0.10 level, after adjusting for multiplicity using the Haseman-Lin-Rahman rules no tests in mice would be categorized as statistical significant. Complete tables of tumor incidence are given in Tables A.2.5 and A.2.6, in Appendix 2, below.

**Table A.2.3. Incidence and Results for Organ-Tumor Combinations in Male Rats**

Organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
<b>Abdominal cavity</b>								
# Evaluated	60	59	60	60				
Adj. # at Risk	45.8	49.1	48.1	46.0				
FIBROMA	0	0	0	1	.2406	.5000	.	.
Adj. # at Risk	45.8	49.1	48.1	45.9				
LIPOMA	0	0	0	1	.2406	.5000	.	.
<b>Adrenal</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.7				
ADENOMA,CORTICAL CELL	0	1	1	1	.2939	.5055	.5161	.5263
Adj. # at Risk	46.4	50.7	48.8	47.5				
PHEOCHROMOCYTOMA	16	11	5	7	.9710	.9936	.9992	.9475
Adj. # at Risk	45.8	50.1	48.1	46.0				
PHEOCHROMOCYTOMA,MALIGNANT	0	1	1	1	.2939	.5055	.5161	.5263
Adj. # at Risk	46.4	50.7	48.8	47.5				
Pheocromocytoma, any	16	11	6	7	.9712	.9936	.9977	.9475
<b>Bone,vertebral</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.7				
CHONDROSARCOMA	0	0	1	1	.1825	.5055	.5161	.
Adj. # at Risk	45.8	50.1	48.1	45.9				
OSTEOSARCOMA	1	0	0	0	1	1	1	1
<b>Cerebellum</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.7	45.9				
GRANULAR CELL TUMOR	0	1	1	0	.6231	.	.5161	.5263
<b>Cerebrum</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.7				
ASTROCYTOMA,MALIGNANT	0	0	0	1	.2434	.5055	.	.
Adj. # at Risk	45.8	50.1	48.1	45.9				
OLIGODENDROGLIOMA	1	0	0	0	1	1	1	1
<b>Forelimb</b>								
# Evaluated	59	59	60	60				
Adj. # at Risk	45.5	49.2	48.1	45.9				
PAPILLOMA,SQUAMOUS CELL	1	0	0	0	1	1	1	1
Adj. # at Risk	45.8	49.2	48.1	45.9				
SCHWANNOMA,MALIGNANT	1	0	0	0	1	1	1	1
Endo. Schwannoma, any	2	0	0	0	1	1	1	1
Adj. # at Risk	45.8	50.1	48.1	45.9				
SCHWANNOMA,ENDOCARDIAL	1	0	0	0	1	1	1	1
Adj. # at Risk	46.5	50.1	48.1	45.9				
SCHWANNOMA,ENDOCARDIAL, MALIGNANT	1	0	0	0	1	1	1	1

**Table A.2.3. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Rats**

Organ/tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
	Veh	Low	Med	High				
Hemolymphoreticular(all sites)								
# Evaluated	60	60	60	60				
Adj. # at Risk	46.8	50.3	48.1	45.9				
LEUKEMIA,LARGE GRANULAR LYMPHO	2	1	0	1	.6569	.8750	1	.8938
Adj. # at Risk	45.8	51.1	48.1	45.9				
LYMPHOMA,MALIGNANT	1	2	0	0	.9416	1	1	.5473
Adj. # at Risk	46.4	50.6	48.1	46.8				
SARCOMA,HISTIOCYTIC	3	3	2	2	.6771	.8195	.8323	.7003
Intestine,ileum								
# Evaluated	57	59	56	57				
Adj. # at Risk	44.2	49.3	45.8	44.2				
ADENOMA	0	0	1	0	.4890	.	.5056	.
Kidney								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
LIPOMA	0	2	0	0	.8157	.	.	.2744
Adj. # at Risk	45.8	50.1	48.1	45.9				
LIPOSARCOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	45.8	50.1	48.1	45.9				
Lipoma/Liposarcoma	1	2	0	0	.9416	1	1	.5399
Liver								
# Evaluated	60	60	60	60				
Adj. # at Risk	46.1	50.1	48.1	45.9				
ADENOMA,HEPATOCELLULAR	5	4	2	0	.9960	1	.9508	.7969
Adj. # at Risk	45.8	50.1	48.1	45.9				
CARCINOMA,HEPATOCELLULAR	0	0	0	1	.2394	.5000	.	.
Adj. # at Risk	45.8	50.1	48.1	45.9				
LYMPHANGIOMA	0	1	0	0	.7606	.	.	.5263
Lung(bronchus)								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
ADENOMA,BRONCHIOLO-ALVEOLAR	0	0	1	0	.4947	.	.5161	.
Mammary gland								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
FIBROADENOMA	1	1	0	0	.9437	1	1	.7783
Adj. # at Risk	45.8	50.1	48.1	46.9				
Fibroadenoma/mixed tumor	1	1	0	1	.5280	.7582	1	.7783
Adj. # at Risk	45.8	50.1	48.1	46.9				
TUMOR,MIXED,BENIGN	0	0	0	1	.2434	.5055	.	.
Mesothelium								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
MESOTHELIOMA,MALIGNANT	1	0	0	1	.4224	.7528	1	1

**Table A.2.3. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Rats**

Organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Oral cavity								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.6	45.9				
CARCINOMA,SQUAMOUS CELL	0	0	1	0	.4947	.	.5161	.
Origin unknown								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
CHORDOMA	0	0	1	0	.4947	.	.5161	.
Palate								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	51.0	48.1	45.9				
CARCINOMA,SQUAMOUS CELL	0	1	0	0	.7606	.	.	.5263
Pancreas								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.4	48.1	46.4				
ADENOMA,ACINAR CELL	1	3	3	1	.6920	.7582	.3332	.3494
Adj. # at Risk	45.8	50.1	48.1	45.9				
ADENOMA,ACINAR-ISLET CELL	1	0	1	0	.7460	1	.7686	1
Adj. # at Risk	47.5	50.4	49.0	45.9				
ADENOMA,ISLET CELL	22	16	10	6	.9997	.9999	.9986	.9558
Adj. # at Risk	45.8	50.4	48.2	46.4				
Acinar Cell Adenoma/Carcin.	1	3	4	1	.7069	.7582	.2014	.3494
Adj. # at Risk	45.8	50.1	48.2	45.9				
CARCINOMA,ACINAR CELL	0	0	1	0	.4947	.	.5161	.
Adj. # at Risk	45.8	50.1	48.1	45.9				
CARCINOMA,ISLET CELL	2	6	2	1	.9075	.8792	.7157	.1708
Adj. # at Risk	47.5	50.4	49.0	45.9				
Islet Cell Adenoma/Carcin.	22	16	11	6	.9997	.9999	.9970	.9558
Parathyroid								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.2	48.1	45.9				
ADENOMA	2	5	0	1	.8932	.8792	1	.2636
Pituitary								
# Evaluated	60	60	60	60				
Adj. # at Risk	49.2	56.3	52.3	48.6				
ADENOMA,ANTERIOR	20	40	23	22	.8277	.3846	.4423	.0014
Adj. # at Risk	45.8	50.1	48.1	45.9				
ADENOMA,INTERMEDIATED	1	0	0	1	.4224	.7528	1	1
Adj. # at Risk	46.4	50.1	48.5	46.1				
CARCINOMA,ANTERIOR	2	0	1	1	.5656	.8791	.8868	1
Prostate								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.0				
ADENOMA	0	1	0	1	.3120	.5055	.	.5263

**Table A.2.3. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Rats**

Organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Skin/Subcutis								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.0				
CARCINOMA,SEBACEOUS	0	0	0	1	.2434	.5055	.	.
Adj. # at Risk	45.8	50.1	48.3	45.9				
CARCINOMA,SQUAMOUS CELL	0	0	1	0	.4947	.	.5161	.
Adj. # at Risk	46.0	50.3	48.1	45.9				
FIBROMA	5	4	2	4	.5297	.7464	.9508	.7969
Adj. # at Risk	45.9	50.4	48.1	45.9				
FIBROSARCOMA	1	1	0	1	.5213	.7528	1	.7783
Adj. # at Risk	46.1	50.1	48.1	46.0				
KERATOACANTHOMA	1	1	1	3	.1054	.3083	.7632	.7730
Adj. # at Risk	45.8	50.1	48.9	45.9				
LEIOMYOSARCOMA	0	0	1	0	.4947	.	.5161	.
Adj. # at Risk	45.8	50.1	48.1	45.9				
LIPOMA	0	0	1	0	.4947	.	.5161	.
Adj. # at Risk	45.8	50.1	48.5	45.9				
LIPOSARCOMA	0	0	1	1	.1792	.5000	.5161	.
Adj. # at Risk	45.8	50.1	48.1	45.9				
PAPILLOMA,SQUAMOUS CELL	1	1	1	0	.8294	1	.7686	.7783
Adj. # at Risk	46.0	50.1	48.3	45.9				
SCHWANNOMA,MALIGNANT	1	0	1	0	.7433	1	.7632	1
Adj. # at Risk	45.8	50.1	48.1	46.0				
TUMOR,HAIR FOLLICLE,BENIGN	1	2	1	1	.5992	.7528	.7686	.5399
Spleen								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
HEMANGIOSARCOMA	0	1	0	0	.7606	.	.	.5263
Adj. # at Risk	45.8	50.1	48.5	45.9				
SARCOMA,NOS	0	0	1	0	.4947	.	.5161	.
Stomach								
# Evaluated	60	60	60	60				
Adj. # at Risk	46.2	50.1	48.1	45.9				
PAPILLOMA,SQUAMOUS CELL	1	1	0	0	.9417	1	1	.7730
Systemic								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
HEMANGIOSARCOMA	0	1	0	0	.7606	.	.	.5263
Testis								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	46.3				
LEYDIG CELL TUMOR	2	0	2	5	.0226	.2264	.7157	1
Adj. # at Risk	45.8	50.1	48.1	45.9				
SEMINOMA	0	1	0	0	.7606	.	.	.5263

**Table A.2.3. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Rats**

Organ/tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
	Veh	Low	Med	High				
Thymus								
# Evaluated	60	60	58	59				
Adj. # at Risk	45.8	50.4	46.7	45.0				
THYMOMA	0	1	0	0	.7581	.	.	.5263
Adj. # at Risk	45.8	50.4	47.6	45.0				
THYMOMA ([B]+[M])	0	1	1	0	.6216	.	.5109	.5263
Adj. # at Risk	45.8	50.1	47.6	45.0				
THYMOMA,MALIGNANT	0	0	1	0	.4920	.	.5109	.
Thyroid								
# Evaluated	58	57	56	59				
Adj. # at Risk	44.9	47.5	46.8	45.7				
ADENOMA,C-CELL	5	5	4	4	.6449	.7691	.7796	.6720
Adj. # at Risk	45.0	47.4	46.4	45.4				
ADENOMA,FOLLICULAR CELL	3	1	1	0	.9725	1	.9469	.9492
Adj. # at Risk	44.9	47.5	46.8	45.9				
C-Cell Adenoma/Carcinoma	5	5	6	6	.3610	.5161	.5319	.6720
Adj. # at Risk	44.9	47.4	46.4	45.6				
CARCINOMA,C-CELL	0	0	2	2	.0705	.2528	.2584	.
Trigeminal nerve								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.8	45.9				
SCHWANNOMA,MALIGNANT	0	0	1	0	.4947	.	.5161	.
Urinary bladder								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
PAPILLOMA,TRANSITIONAL CELL	0	0	1	0	.4947	.	.5161	.
Zymbal gland								
# Evaluated	60	60	60	60				
Adj. # at Risk	45.8	50.1	48.1	45.9				
ADENOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	46.4	50.1	48.5	45.9				
Adenoma/Carcinoma	2	0	2	0	.8206	1	.7076	1
Adj. # at Risk	46.4	50.1	48.5	45.9				
CARCINOMA	1	0	2	0	.6733	1	.5161	1

**Table A.2.4. Incidence and Results for Organ-Tumor Combinations in Female Rats**

Organ/tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
<b>Adrenal</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.5	41.5	42.4	49.4				
ADENOMA,CORTICAL CELL	1	1	2	1	.5562	.7843	.4911	.7410
Adj. # at Risk	43.3	41.4	42.7	49.4				
CARCINOMA,CORTICAL CELL	0	0	1	0	.5200	.	.4941	.
Adj. # at Risk	43.5	41.5	42.7	49.4				
Cortical Adenoma/Carcinoma	1	1	3	1	.5917	.7843	.2991	.7410
Adj. # at Risk	43.3	41.4	42.4	49.5				
PHEOCHROMOCYTOMA	0	0	1	3	.0261	.1467	.4941	.
Adj. # at Risk	43.3	41.4	42.4	49.4				
PHEOCHROMOCYTOMA,MALIGNANT	0	1	0	0	.7543	.	.	.4881
Adj. # at Risk	43.3	41.4	42.4	49.5				
Pheochromocytoma, Any	0	1	1	3	.0590	.1467	.4941	.4881
<b>Cerebrum</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.4				
ASTROCYTOMA,MALIGNANT	1	0	1	1	.4347	.7843	.7471	1
Adj. # at Risk	43.3	41.4	42.4	49.4				
PAPILLOMA,CHOROID PLEXUS	0	1	0	0	.7543	.	.	.4881
<b>Hemolymphoreticular(all sites)</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	44.1	41.4	42.4	49.4				
LEUKEMIA,LARGE GRANULAR LYMPHO	1	0	0	0	1	1	1	1
Adj. # at Risk	43.3	41.6	42.4	49.9				
LYMPHOMA,MALIGNANT	0	1	0	1	.3444	.5326	.	.4881
Adj. # at Risk	43.6	41.4	42.7	49.4				
SARCOMA,HISTIOCYTIC	2	0	1	0	.9036	1	.8751	1
<b>Intestine, jejunum</b>								
# Evaluated	58	56	57	58				
Adj. # at Risk	41.5	39.9	41.6	48.8				
LEIOMYOSARCOMA	0	1	0	0	.7574	.	.	.4875
<b>Kidney</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.4				
ADENOMA,RENAL CELL	0	1	0	0	.7543	.	.	.4881
Adj. # at Risk	43.3	41.4	43.2	49.4				
LIPOMA	0	0	1	0	.5227	.	.5000	.
<b>Liver</b>								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.9				
ADENOMA,HEPATOCELLULAR	1	0	0	1	.4828	.7843	1	1

**Table A.2.4. (cont.) Incidence and Results for Organ-Tumor Combinations in Female Rats**

Organ/tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	phigh	pmed	plow
	Veh	Low	Med	High				
Mammary gland								
# Evaluated	60	60	60	60				
Adj. # at Risk	47.7	47.0	48.0	52.9				
ADENOCARCINOMA	23	18	18	18	.8704	.9506	.9088	.8941
Adj. # at Risk	45.2	44.1	43.2	49.8				
ADENOCARCINOMA IN FIBROADENOMA	6	9	2	4	.9045	.8743	.9663	.2700
Adj. # at Risk	43.3	42.5	42.8	49.4				
ADENOMA	0	3	2	1	.6138	.5326	.2412	.1162
Adj. # at Risk	49.2	49.0	50.5	53.1				
Adenoma/Fibro--/carc/carc in	36	28	32	28	.9594	.9912	.8913	.9634
Adj. # at Risk	45.3	44.0	45.9	50.1				
FIBROADENOMA	20	14	15	15	.8651	.9527	.9029	.9139
Adj. # at Risk	43.3	41.4	42.4	49.4				
LIPOMA	0	1	0	0	.7543	.	.	.4881
Adj. # at Risk	43.8	41.4	42.4	50.2				
TUMOR,MIXED,MALIGNANT	2	0	0	1	.6356	.9049	1	1
Oral cavity								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.5	43.0	49.7				
CARCINOMA,SQUAMOUS CELL	0	1	1	1	.3339	.5326	.4941	.4881
Ovary								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.3	41.4	42.4	49.4				
GRANULOSA CELL TUMOR	0	0	1	0	.5200	.	.4941	.
Adj. # at Risk	43.3	41.4	42.9	49.4				
THECOMA,BENIGN	0	0	1	0	.5200	.	.4941	.
Pancreas								
# Evaluated	60	60	60	60				
Adj. # at Risk	43.9	41.6	42.4	49.4				
ADENOMA,ISLET CELL	5	4	4	2	.9204	.9622	.7465	.7340
Adj. # at Risk	43.3	41.4	42.4	49.4				
CARCINOMA,ACINAR CELL	0	0	1	0	.5200	.	.4941	.
Adj. # at Risk	43.3	41.4	42.4	49.4				
CARCINOMA,ISLET CELL	0	0	0	1	.2800	.5326	.	.
Adj. # at Risk	43.9	41.6	42.4	49.4				
Islet Cell Adenoma/Carcinoma	5	4	4	3	.8239	.9045	.7465	.7340
Parathyroid								
# Evaluated	58	59	59	59				
Adj. # at Risk	41.9	41.3	41.9	49.4				
ADENOMA	1	0	2	0	.7303	1	.5000	1
Pituitary								
# Evaluated	60	60	60	60				
Adj. # at Risk	58.1	56.7	55.4	54.7				
ADENOMA,ANTERIOR	49	44	44	38	.9572	.9787	.8079	.8544
Adj. # at Risk	58.7	58.0	56.4	55.5				
Anterior Adenoma/Carcinoma	51	48	47	41	.9633	.9814	.8116	.8532
Adj. # at Risk	43.9	42.7	43.4	50.3				
CARCINOMA,ANTERIOR	2	4	3	3	.5633	.5719	.5000	.3267



**Table A.2.4. (cont.) Incidence and Results for Organ-Tumor Combinations in Female Rats**

Organ/tumor	Overall Results				Significance Levels					
	Tumor Incidence				ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh		
	Veh	Low	Med	High						
Zymbal gland										
# Evaluated	60	60	60	60						
Adj. # at Risk	43.3	41.4	42.4	49.5						
CARCINOMA	1	0	0	1	.4828	.7843	1	1	1	1
Adj. # at Risk	43.3	41.4	42.4	49.5						
PAPILLOMA,SQUAMOUS CELL	0	0	0	1	.2800	.5326	.	.		
Adj. # at Risk	43.3	41.5	42.4	49.5						
Sq.Cell Carc,Pap,Keratoacanthoma	0	2	0	1	.4805	.5326	.		.2352	

The following two tables give similar results in mice. Again, for each identified neoplasm within organ, the adjusted number at risk is presented first. The next row provides the tumor incidence over all five dose groups, followed by the significance levels of test of trend over the actual dose groups 1-4, and then followed by the results of the comparisons between the high dose and the high-medium dose, respectively, with the vehicle. The next row, with slightly indented p-values lined up with those of the preceding row, presents the significance levels of the comparisons between the low and positive control, respectively, with vehicle.

**Table A.2.5. Incidence and Results for Organ-Tumor Combinations in Male Mice**

Organ/Tumor	Overall Results				Significance Levels					
	Tumor Incidence				ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh		
	Veh	Low	Med	High						
Adrenal										
# Evaluated	55	55	55	55						
Adj. # at Risk	42.4	51.5	52.0	49.0						
ADENOMA,SUBCAPSULAR CELL	2	0	0	0	1	1	1	1		
Cranial bone										
# Evaluated	55	54	55	55						
Adj. # at Risk	42.4	50.6	52.7	49.0						
OSTEOSARCOMA	0	0	1	0	.5208	.	.5532	.		
Harderian gland										
# Evaluated	55	55	55	55						
Adj. # at Risk	42.4	51.5	52.0	49.0						
ADENOCARCINOMA	0	0	0	1	.2500	.5333	.	.		
Adj. # at Risk	43.1	51.5	52.0	49.0						
ADENOMA	5	7	5	3	.8807	.8992	.7334	.5053		
Adj. # at Risk	42.4	51.5	52.0	49.0						
Adenoma/Adenocarcinoma	0	0	0	1	.2500	.5333	.	.		
Heart										
# Evaluated	55	55	55	55						
Adj. # at Risk	42.4	51.5	52.0	49.0						
HEMANGIOSARCOMA	1	0	0	0	1	1	1	1		

**Table A.2.5. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Mice**

Organ/Tumor	Overall Results Tumor Incidence				Significance Levels			
	Veh	Low	Med	High	ptrend	p <sub>high</sub> vsVeh	p <sub>med</sub> vsVeh	p <sub>low</sub> vsVeh
Hemolymphoreticular(all sites)								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.6	52.0	52.7	49.0				
LYMPHOMA,MALIGNANT	7	7	5	5	.8373	.8811	.9079	.7664
Adj. # at Risk	43.6	51.5	52.0	49.1				
SARCOMA,HISTIOCYTIC	3	1	0	2	.7323	.8575	1	.9595
Intestine, overall								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA	1	1	0	0	.9530	1	1	.7987
Intestine,cecum								
# Evaluated	51	53	53	47				
Adj. # at Risk	40.2	50.3	50.7	45.7				
LEIOMYOSARCOMA	0	1	0	0	.7838	.	.	.5556
Intestine,duodenum								
# Evaluated	52	51	51	47				
Adj. # at Risk	40.6	48.7	49.7	45.7				
ADENOMA	1	0	0	0	1	1	1	1
Intestine,ileum								
# Evaluated	52	52	52	50				
Adj. # at Risk	40.3	49.3	49.8	47.2				
ADENOMA	0	1	0	0	.7838	.	.	.5506
Adj. # at Risk	40.8	49.3	49.8	47.2				
HEMANGIOMA	1	0	0	0	1	1	1	1
Kidney								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA,RENAL CELL	0	0	0	1	.2500	.5333	.	.
Liver								
# Evaluated	55	55	55	55				
Adj. # at Risk	43.7	52.3	52.0	49.1				
ADENOMA,HEPATOCELLULAR	15	10	7	7	.9897	.9950	.9962	.9749
Adj. # at Risk	44.0	52.5	52.6	49.6				
CARCINOMA,HEPATOCELLULAR	9	7	8	5	.9086	.9578	.8342	.8930
Adj. # at Risk	42.4	51.5	52.0	49.0				
CARCINOMA,HEPATOCHOLANGIO-CELLULAR	0	1	0	0	.7812	.	.	.5484
Adj. # at Risk	42.4	51.5	52.0	49.0				
HEMANGIOMA	0	0	1	0	.5156	.	.5484	.
Adj. # at Risk	42.4	51.6	52.0	49.0				
HEMANGIOSARCOMA	1	1	1	0	.8684	1	.7987	.7987
Adj. # at Risk	45.3	52.9	52.6	49.7				
Hepato. Adenoma/Carc./Hemangioma	22	16	15	12	.9897	.9964	.9875	.9791

**Table A.2.5. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Mice**

Organ/Tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	p <sub>high</sub>	p <sub>med</sub>	p <sub>low</sub>
	Veh	Low	Med	High	vsVeh	vsVeh	vsVeh	vsVeh
Lung(bronchus)								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.6	51.5	52.0	49.1				
ADENOMA, BRONCHIOLO-ALVEOLAR	5	2	3	4	.6298	.8283	.9238	.9689
Adj. # at Risk	42.5	51.5	52.0	49.0				
CARCINOMA, BRONCHIOLO-ALVEOLAR	2	1	1	0	.9566	1	.9115	.9115
Pancreas								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA, ACINAR CELL	0	0	0	1	.2500	.5333	.	.
Skin+Subcutis								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
HEMANGIOMA	0	1	0	0	.7812	.	.	.5484
Adj. # at Risk	42.4	51.5	52.0	49.0				
HEMANGIOSARCOMA	0	0	1	0	.5156	.	.5484	.
Adj. # at Risk	42.4	51.5	52.0	49.0				
Hemangioma/-sarcoma	0	1	1	0	.6499	.	.5484	.5484
Adj. # at Risk	42.4	51.5	52.0	49.8				
LEIOMYSARCOMA	0	0	0	1	.2539	.5385	.	.
Systemic								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.8	51.5	52.0	49.0				
HEMANGIOMA	1	1	1	0	.8684	1	.7987	.7987
Adj. # at Risk	42.4	51.6	52.0	49.0				
HEMANGIOSARCOMA	2	1	2	0	.9219	1	.7611	.9115
Adj. # at Risk	42.8	51.6	52.0	49.0				
Hemangioma/-sarcoma	3	2	3	0	.9615	1	.7490	.8738
Testis								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
LEYDIG CELL TUMOR	0	0	1	2	.0644	.2816	.5484	.
Thymus								
# Evaluated	54	55	54	53				
Adj. # at Risk	42.1	51.5	51.0	47.5				
THYMOMA	0	0	1	0	.5105	.	.5435	.
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
ADENOMA, FOLLICULAR CELL	0	0	1	2	.0644	.2816	.5484	.
Adj. # at Risk	42.4	51.5	52.0	49.0				
CARCINOMA, FOLLICULAR CELL	0	0	1	0	.5181	.	.5532	.
Adj. # at Risk	42.4	51.5	52.0	49.0				
Foll. Cell Adenoma/Carcinoma	0	0	2	2	.0749	.2816	.3034	.

**Table A.2.5. (cont.) Incidence and Results for Organ-Tumor Combinations in Male Mice**

Organ/Tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	phigh	pmed	plow
	Veh	Low	Med	High				
Tongue								
# Evaluated	55	55	55	55				
Adj. # at Risk	42.4	51.5	52.0	49.0				
PAPILLOMA, SQUAMOUS CELL	1	0	0	0	1	1	1	1

**Table A.2.6. Incidence and Results for Organ-Tumor Combinations in FemMale Mice**

Organ/Tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	phigh	pmed	plow
	Veh	Low	Med	High				
Adrenal								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOMA, CORTICAL CELL	0	0	1	0	.4721	.	.5000	.
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOMA, SUBCAPSULAR CELL	1	1	0	0	.9313	1	1	.7524
Adj. # at Risk	52.4	52.8	52.1	41.1				
PHEOCHROMOCYTOMA	0	2	0	1	.3914	.4409	.	.2476
Adj. # at Risk	52.4	52.8	52.1	41.1				
PHEOCHROMOCYTOMA, MALIGNANT	0	1	0	0	.7360	.	.	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
Pheochromocytoma, Any	0	3	0	1	.4791	.4409	.	.1214
Bone+Bone marrow, femoral								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
HEMANGIOSARCOMA	0	1	0	0	.7360	.	.	.5000
Brachium								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
SARCOMA, NOS	0	0	1	0	.4721	.	.5000	.
Buccal								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
OSTEOSARCOMA	0	1	0	0	.7360	.	.	.5000
Harderian gland								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.4	41.6				
ADENOMA	5	6	2	3	.7913	.7748	.9439	.5000
Hemolymphoreticular(all sites)								
# Evaluated	55	55	55	55				
Adj. # at Risk	53.6	53.2	52.5	41.1				
LYMPHOMA, MALIGNANT	24	20	15	5	.9999	.9999	.9745	.8378
Adj. # at Risk	52.7	52.8	53.2	41.1				
SARCOMA, HISTIOCYTIC	4	0	7	1	.6544	.9500	.2741	1

**Table A.2.6. (cont.) Incidence and Results for Organ-Tumor Combinations in Female Mice**

organ/tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	phigh	pmed	plow
	Veh	Low	Med	High				
Intestine, duodenum								
# Evaluated	52	55	50	52				
Adj. # at Risk	50.3	52.8	48.1	39.3				
ADENOMA	0	1	0	0	.7354	.	.	.5098
Kidney								
# Evaluated	55	55	54	55				
Adj. # at Risk	52.4	52.8	51.4	41.1				
ADENOMA, RENAL CELL	1	0	0	0	1	1	1	1
Adj. # at Risk	52.4	52.8	51.4	41.1				
CARCINOMA, RENAL CELL	2	0	0	0	1	1	1	1
Adj. # at Risk	52.4	52.8	51.4	41.1				
Renal Cell Adenoma/carcinoma	2	0	0	0	1	1	1	1
Liver								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.3				
ADENOMA, HEPATOCELLULAR	12	6	3	3	.9934	.9923	.9980	.9662
Adj. # at Risk	52.6	52.8	52.1	41.1				
CARCINOMA, HEPATOCELLULAR	3	4	6	1	.7298	.9073	.2439	.5000
Adj. # at Risk	52.7	52.8	52.1	41.1				
HEMANGIOSARCOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	52.6	52.8	52.1	41.3				
Hepato. Adenoma/carcinoma	15	9	9	4	.9897	.9955	.9489	.9489
Lung (bronchus)								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.6	52.8	52.1	41.1				
ADENOMA, BRONCHIOLO-ALVEOLAR	3	1	1	1	.8227	.9073	.9411	.9411
Adj. # at Risk	52.6	52.8	52.1	41.1				
Bronch-Alv. Adenoma/Carc.	6	2	2	2	.8971	.9388	.9701	.9701
Adj. # at Risk	52.4	52.8	52.1	41.1				
CARCINOMA, BRONCHIOLO-ALVEOLAR	3	1	1	1	.8227	.9073	.9411	.9411
Mammary gland								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOACANTHOMA, MALIGNANT	0	2	0	0	.7913	.	.	.2476
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOCARCINOMA	0	2	0	0	.7913	.	.	.2476
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOMA	1	1	0	0	.9313	1	1	.7524
Adj. # at Risk	52.4	52.8	52.1	41.1				
Adenoma/-carc./-canthoma	1	4	0	0	.9365	1	1	.1813

**Table A.2.6. (cont.) Incidence and Results for Organ-Tumor Combinations in Female Mice**

organ/tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	phigh	pmed	plow
	Veh	Low	Med	High				
<b>Ovary</b>								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
CYSTADENOMA	0	1	0	0	.7360	.	.	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
GRANULOSA CELL TUMOR	0	0	1	0	.4721	.	.5000	.
Adj. # at Risk	52.4	52.8	52.3	41.1				
HEMANGIOMA	0	1	1	0	.5825	.	.5000	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
LUTEOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	52.4	52.8	52.1	41.1				
SERTOLI CELL TUMOR	1	1	0	0	.9313	1	1	.7524
Adj. # at Risk	52.4	52.8	52.6	41.1				
TERATOMA	0	0	1	0	.4721	.	.5000	.
Adj. # at Risk	52.4	52.8	52.1	41.1				
YOLK SAC CARCINOMA	0	1	0	0	.7360	.	.	.5000
<b>Pancreas</b>								
# Evaluated	55	55	54	55				
Adj. # at Risk	52.4	52.8	51.4	41.3				
ADENOMA, ISLET CELL	0	0	0	1	.2092	.4409	.	.
<b>Pituitary</b>								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.9	41.6				
ADENOMA, ANTERIOR	10	8	10	2	.9701	.9936	.5980	.7812
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOMA, INTERMEDIATED	0	1	1	0	.5825	.	.5000	.5000
<b>Skin+Subcutis</b>								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	53.2	52.1	41.1				
CARCINOMA, BASAL CELL	0	1	0	0	.7374	.	.	.5048
Adj. # at Risk	52.4	52.8	52.1	41.1				
FIBROMA	1	0	0	0	1	1	1	1
Adj. # at Risk	52.4	52.8	52.2	41.6				
FIBROSARCOMA	1	0	2	1	.3506	.6900	.5000	1
Adj. # at Risk	52.4	52.8	52.1	41.1				
HEMANGIOSARCOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	52.4	53.2	52.1	41.1				
LIPOSARCOMA	1	1	0	0	.9320	1	1	.7571
Adj. # at Risk	52.4	52.8	52.1	41.1				
MELANOMA, MALIGNANT	0	1	1	0	.5825	.	.5000	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
SARCOMA, NOS	0	1	1	0	.5825	.	.5000	.5000
<b>Spleen</b>								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
HEMANGIOMA	1	0	0	0	1	1	1	1

**Table A.2.6. (cont.) Incidence and Results for Organ-Tumor Combinations in Female Mice**

organ/tumor	Overall Results				Significance Levels			
	Tumor Incidence				ptrend	phigh	pmed	plow
	Veh	Low	Med	High				
Stomach								
# Evaluated	54	55	54	54				
Adj. # at Risk	51.7	52.8	51.1	40.3				
NEUROENDOCRINE TUMOR, MALIG-NANT	1	0	0	0	1	1	1	1
Adj. # at Risk	51.7	52.8	51.1	40.5				
PAPILLOMA, SQUAMOUS CELL	1	0	1	2	.1498	.4088	.7525	1
Systemic								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.3	41.1				
HEMANGIOMA	1	1	2	0	.7440	1	.5000	.7524
Adj. # at Risk	52.7	52.8	52.1	41.1				
HEMANGIOSARCOMA	2	2	1	0	.9365	1	.8786	.6912
Adj. # at Risk	52.7	52.8	52.3	41.1				
Hemangioma/-sarcoma	3	3	3	0	.9338	1	.6609	.6609
Thoracic cavity								
# Evaluated	53	54	55	55				
Adj. # at Risk	50.6	51.9	52.7	41.1				
OSTEOSARCOMA	0	0	1	0	.4794	.	.5098	.
Thyroid								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.2	41.1				
ADENOMA, FOLLICULAR CELL	1	1	1	3	.0889	.2245	.7524	.7524
Uterus								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
ADENOCARCINOMA	1	2	1	0	.8511	1	.7524	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
CYSTADENOMA	0	1	0	0	.7360	.	.	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
GRANULAR CELL TUMOR	0	1	1	0	.5825	.	.5000	.5000
Adj. # at Risk	52.4	52.8	52.1	41.1				
HEMANGIOMA	0	0	1	0	.4721	.	.5000	.
Adj. # at Risk	52.4	52.8	52.1	41.1				
HEMANGIOSARCOMA	0	2	0	0	.7913	.	.	.2476
Adj. # at Risk	52.4	52.8	52.1	41.1				
LEIOMYOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	52.4	52.8	52.1	41.1				
LEIOMYOSARCOMA	0	0	1	0	.4721	.	.5000	.
Adj. # at Risk	52.4	52.8	52.1	42.0				
POLYP, ENDOMETRIAL STROMAL	2	4	1	3	.3897	.3989	.8786	.3391
Vagina								
# Evaluated	55	55	55	55				
Adj. # at Risk	52.4	52.8	52.1	41.1				
HEMANGIOSARCOMA	0	0	1	0	.4721	.	.5000	.

### Appendix 3. References

- Bailer, A. and Portier, C. (1988), "Effects of Treatment-Induced Mortality on Tests for Carcinogenicity in Small Samples", *Biometrics*, **44**, 4, 417-431.
- Bieler, G.S., and Williams, R.L. (1993), "Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity", *Biometrics*, **49**, 4, 793-801.
- Chu, K.C., Ceuto, C., and Ward, J.M. (1981), "Factors in the Evaluation of 200 National Cancer Institute Carcinogen Bioassays", *Journal of Toxicology and Environmental Health*, **8**, 251-280.
- Greaves, P. (2007), "Neoplasia of Adrenal Medulla," In: *Histopathology of Preclinical Toxicity Studies*. 3<sup>rd</sup> edition, pp 819. Oxford, UK: Academic Press, Elsevier Inc.
- Haseman, J. K. (1983), "A Reexamination of False-positive Rates for Carcinogenicity Studies", *ndamental and Applied Toxicology*, **3**, 334-339.
- Jara, A. (2007), "Applied Bayesian Non- and Semi-parametric Inference using DPpackage", *Rnews*, **7**, 3, 17-26.
- Lin, K. K. and Ali, M.W. (2006), "Statistical Review and Evaluation of Animal Tumorigenicity Studies", *Statistics in the Pharmaceutical Industry, Third Edition*, edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.
- Lin, K. K. and Rahman, M.A. (1998), "Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs", *Journal of Biopharmaceutical Statistics*. **8**, 1, 1-15.
- McConnell, E.E., Solleveld, H.A., Swenberg, J.A., and Boorman, G.A. (1986), "Guidelines for Combining Neoplasms for Evaluation of Rodent Carcinogenesis Studies", *Journal of the National Cancer Institute*. **76**, 283-289.
- Parola, A, and Jacobs, A. (2010). "Combining Tumors for Statistical Analysis", online FDA handout.
- Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). "Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments", *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal*, International Agency for Research Against Cancer, 311-426.

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

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

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

U.S. Department of Health and Human Services (2013), Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (DRAFT GUIDANCE), Center for Drug Evaluation and Research, Food and Drug Administration

---

i

-----  
**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

11/19/2015

Carcinogenicity Statistical Review  
(Late because of archiving delay)

KARL K LIN

11/20/2015

Concur with review



US 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

### Biometrics Division: VI

<b>NDA No.:</b>	207947
<b>DATE RECEIVED BY OB:</b>	June 18, 2015
<b>DRUG NAME:</b>	Selexipag
<b>INDICATION:</b>	Treatment of patients with pulmonary arterial hypertension (PAH)
<b>SPONSOR:</b>	Actelion
<b>REVIEW FINISHED DATE:</b>	August 21, 2015
<b>CMC STATISTICAL REVIEWER:</b>	Zhuang Miao, Ph.D.
<b>CMC REVIEWER:</b>	Mariappan Chelliah, Ph.D.

**Secondary Reviewer:**

Meiyu Shen, Ph.D., Team Leader, Mathematical Statistician, CDER/OTS/OB/DB VI

**Concur:**

Yi Tsong, Ph.D., Division Director, DBVI, CDER/OTS/OB/DB VI

**Distribution:**

CDER/OTS/OB/DB VI/ Yi Tsong

CDER/OTS/OB/DB VI/ Meiyu Shen

CDER/OTS/OB/ Lillian Patrician

CDER/OPQ/ONDP/DNDPI/NDPBI Mariappan Chelliah

CDER/OPQ/OPRO/DRBPMI/RBPMBII Maryam Kord Bacheh Changi

CDER/OPQ/ONDP/DNDPI/NDPBI Wendy Wilson

## Table of Contents

<i>I. Executive summary.....</i>	<i>5</i>
<i>II. Purpose of the review.....</i>	<i>5</i>
<i>III. Sponsor's analysis and results.....</i>	<i>5</i>
<i>Reviewer's Comments on the Sponsor's Analysis:.....</i>	<i>6</i>
<i>IV. FDA's information request and the sponsor's response to FDA IR.....</i>	<i>6</i>
<i>V. FDA Statistical reviewer's Analyses.....</i>	<i>7</i>
V.1 Shelf life estimation for dose strength 200µg.....	9
V.2 Shelf life estimation for dose strength 400µg.....	10
V.3 Shelf life estimation for dose strength 800µg.....	11
V.4 Shelf life estimation for dose strength 1200µg.....	12
V.5 Shelf life estimation for dose strength 1600µg.....	14
<i>VI. Conclusions and recommendations.....</i>	<i>15</i>

**List of Tables:**

Table 1: FDA Statistics Reviewer’s Estimated Shelf Life for each Dose Strength using Long-term Stability Data .....	5
Table 2: Sponsor’s Shelf Life Estimation for Selexipag packed in bottles (worst case).....	6
Table 3: Summary of Data submitted on July 13, 2015 .....	7
Table 4: Summary of Initial Assay (%) of each batch.....	8
Table 5: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 200µg.....	9
Table 6: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 400µg.....	10
Table 7: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 800µg.....	11
Table 8: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 1200µg.....	13
Table 9: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 1600µg.....	14
Table 10: FDA Statistics Reviewer’s Estimated Shelf Life for each Dose Strength using Long-term Stability Data .....	15

**List of Figures:**

Figure 1: Stability Plot of Assay for Dose Strength 200 $\mu$ g under the Long-term Conditions ..... 9  
Figure 2: Stability Plot of Assay for Dose Strength 400 $\mu$ g under the Long-term Conditions ..... 11  
Figure 3: Stability Plot of Assay for Dose Strength 800 $\mu$ g under the Long-term Conditions ..... 12  
Figure 4: Stability Plot of Assay for Dose Strength 1200 $\mu$ g under the Long-term Conditions ... 13  
Figure 5: Stability Plot of Assay for Dose Strength 1600 $\mu$ g under the Long-term Conditions ... 14

## I. EXECUTIVE SUMMARY

From the release and stability data, it is clear that all release values are above (b) (4)% and some of them are very close to the upper limit (b) (4)% and there is a (b) (4) for each dose. Assay will (b) (4) (b) (4)% if future batches are manufactured (b) (4) (b) (4)%. This concern is especially about the dose strength 1200 µg (b) (4) (b) (4).

We performed statistical analysis on (b) (4) months long-term stability data with five dose strengths (200µg, 400µg, 800µg, 1200µg, 1600µg). Each dose strength has three batches. Our estimations on the shelf life for each dose strength are summarized in Table 1 below.

**Table 1: FDA Statistics Reviewer's Estimated Shelf Life for each Dose Strength using Long-term Stability Data**

Does strength	Package	Shelf life estimation in month
200µg	Bottle	(b) (4)
400µg	Bottle	(b) (4)
800µg	Bottle	(b) (4)
1200µg	Bottle	(b) (4)
1600µg	Bottle	(b) (4)

The estimated shelf lives for dose strength 200µg, 400µg, 800µg and 1600µg are longer than (b) (4) months. However, the estimated shelf life for dose strength 1200µg is (b) (4) months. (b) (4)

## II. PURPOSE OF THE REVIEW

On June 18, 2015, Office of New Drug Products requested the CMC statistics team in Office of Biostatistics to evaluate the sponsor's shelf life estimation for NDA 207947. The sponsor proposed a 36 months shelf life based on their statistical analysis using a (b) (4). The ONDP reviewer requested the OB reviewer to conduct the analysis in order to determine if the proposed 36 months shelf life for the new drug is supported.

## III. SPONSOR'S ANALYSIS AND RESULTS

(b) (4)



*Hence the sponsor's model cannot be accepted for predicting the shelf life.*

**IV. FDA'S INFORMATION REQUEST AND THE SPONSOR'S RESPONSE TO FDA IR**

During the review cycle, FDA sent one information request to the sponsor in order to facility the review.

On July 6, 2015, FDA sent out information request below,

“Provide us a copy of the data set in SAS format as well as the SAS code that were used to generate the report in section 3.2.P.8.3 of the eCTD titled ‘Statistical Evaluation of the Registration Stability Package for Selexipag Tablets’.”

On July 13, 2015, the sponsor submitted their response to information request with the data set used for the shelf life estimation. The data summary is in the following table.

**Table 3: Summary of Data submitted on July 13, 2015**

Does strength	Package	Temperature	Replication	Measured time point (month)
200 µg	Bottle	(b) (4)	(b) (4)	(b) (4)
400 µg	Bottle	(b) (4)	(b) (4)	(b) (4)
800 µg	Bottle	(b) (4)	(b) (4)	(b) (4)
1200 µg	Bottle	(b) (4)	(b) (4)	(b) (4)
1600 µg	Bottle	(b) (4)	(b) (4)	(b) (4)

**V. FDA STATISTICAL REVIEWER’S ANALYSES**

We found out that all the initial assays (%) of each batch are above (b) (4)% and some of them are very close to the upper limit (b) (4)%. For example, batch 12-01969 has initial Assay (b) (4)% shown in Table 4. This results in a (b) (4) of the assay value for each

dose.

(b) (4)

**Table 4: Summary of Initial Assay (%) of each batch**

<b>Does strength</b>	<b>Batch Number</b>	<b>Initial Assay (%)</b>
<b>200 µg</b>	12-01962	(b) (4)
	12-01963	
	12-01964	
<b>400 µg</b>	12-01965	
	12-01966	
	12-01967	
<b>800 µg</b>	12-01968	
	12-01969	
	12-01970	
<b>1200 µg</b>	12-01971	
	12-01972	
	12-01973	
<b>1600 µg</b>	12-01974	
	12-01975	
	12-01976	

We performed independent by-dose statistical analysis on the long-term stability data using ANCOVA method according to "Guidance for Industry Q1E Evaluation of Stability Data".

(b) (4)

Based on the input of Dr. Mariappan Chelliah (ONDP), we use one-sided 95% confidence interval and (b) (4)% label claim to estimate the shelf life

(b) (4)

the shelf life is estimated by

(b) (4)

(b) (4)

### V.1 Shelf life estimation for dose strength 200µg

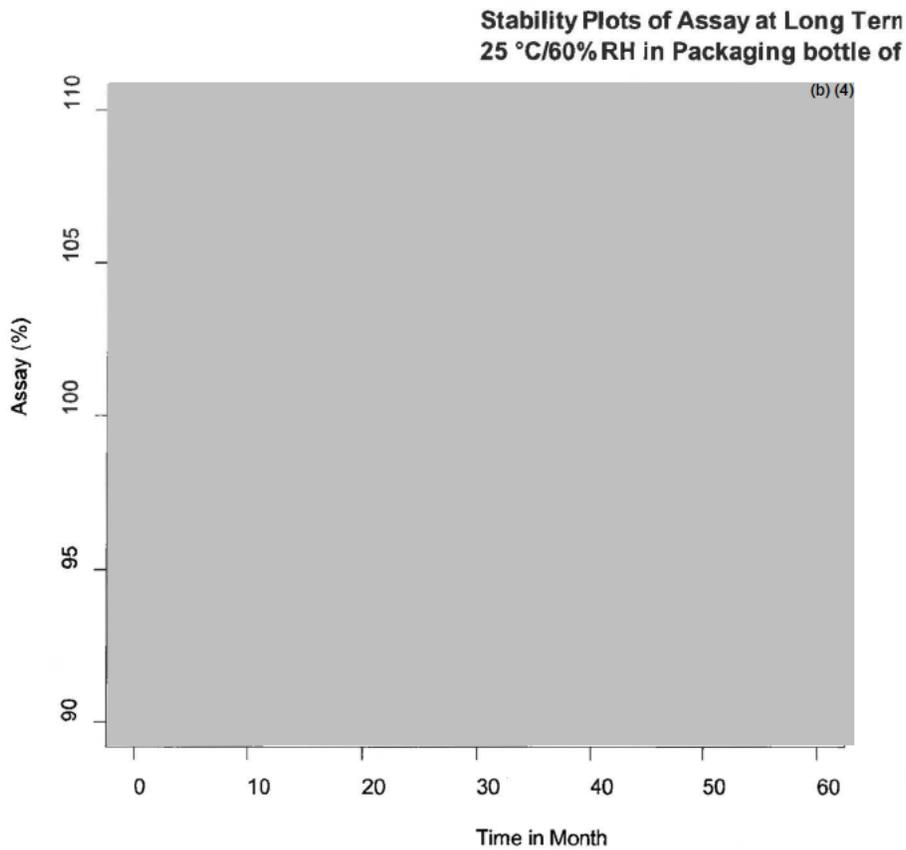
For dose strength 200µg, we performed the poolability test on packaging bottle based on the approach outlined in ICH Q1E guidance.



**Table 5: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 200µg**

The content of Table 5 is redacted with a solid grey fill. A small '(b) (4)' label is located in the top right corner of the redacted area.

**Figure 1: Stability Plot of Assay for Dose Strength 200µg under the Long-term Conditions**



In

[Redacted]

(b) (4)

(b) (4) Thus, the stability analysis supports a shelf life of 36 months for packaging bottle by a common-slope-different-intercept model.

**V.2 Shelf life estimation for dose strength 400µg**

For dose strength 400µg, we performed the poolability test on packaging bottle based on the approach outlined in ICH Q1E guidance.

[Redacted]

(b) (4)

**Table 6: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 400µg**

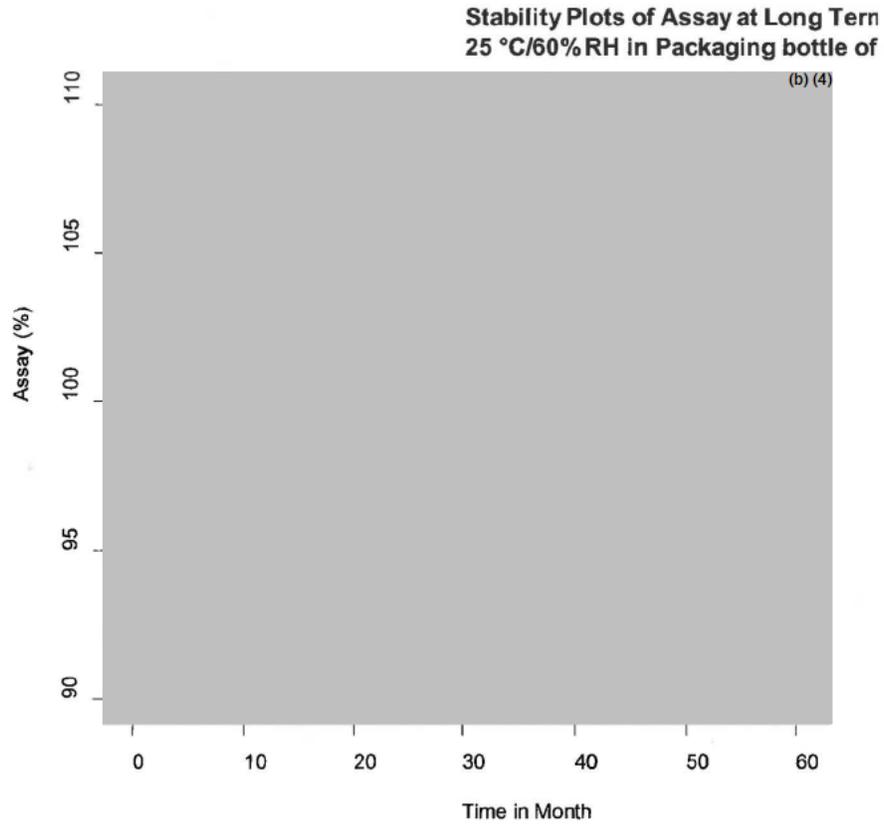
[Redacted Table Content]

(b) (4)

(b) (4)

(b) (4) Thus, the stability analysis supports a shelf life of 36 months for packaging bottle by a common-slope-different-intercept model.

**Figure 2: Stability Plot of Assay for Dose Strength 400µg under the Long-term Conditions**



### V.3 Shelf life estimation for dose strength 800µg

For dose strength 800µg, we performed the poolability test on packaging bottle based on the approach outlined in ICH Q1E guidance.

[Redacted content] (b) (4)

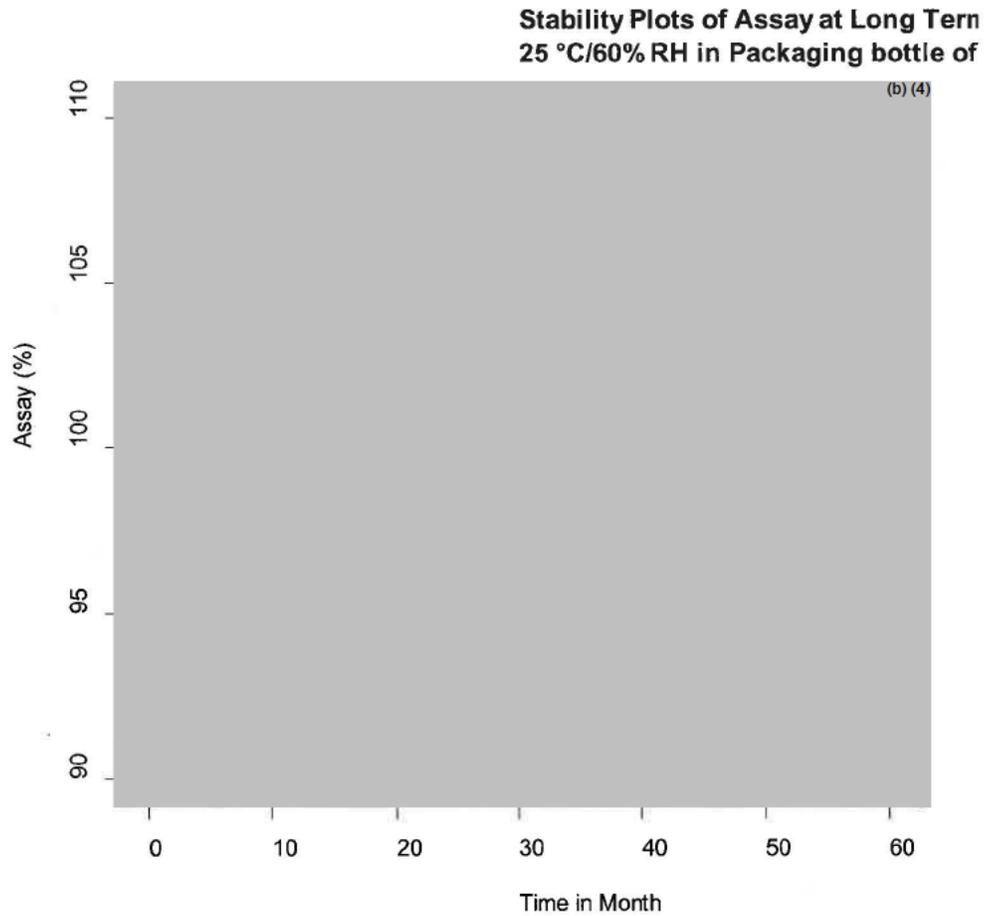
**Table 7: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 800µg**

[Redacted content] (b) (4)

(b) (4)

(b) (4) Thus, the stability analysis supports a shelf life of 36 months for packaging bottle by a common-slope-different-intercept model.

**Figure 3: Stability Plot of Assay for Dose Strength 800µg under the Long-term Conditions**



#### **V.4 Shelf life estimation for dose strength 1200µg**

For dose strength 1200µg, we performed the poolability test on packaging bottle based on the approach outlined in ICH Q1E guidance.

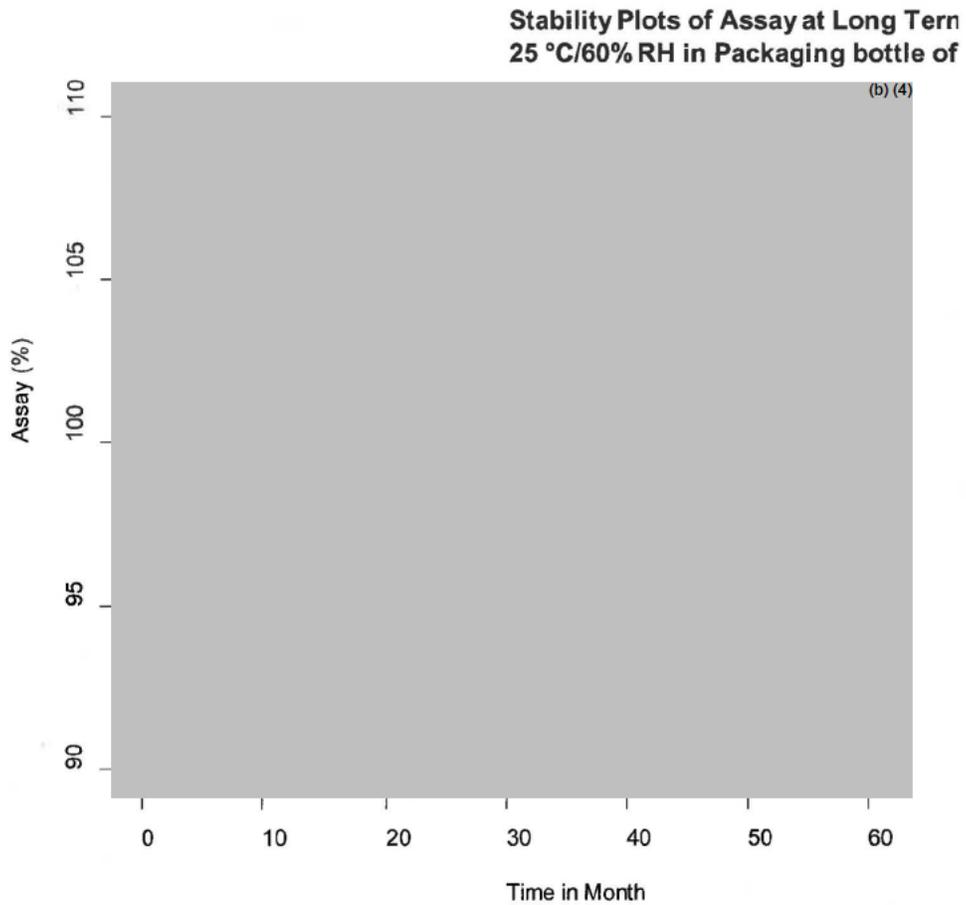
(b) (4)

**Table 8: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 1200µg**

(b) (4)



**Figure 4: Stability Plot of Assay for Dose Strength 1200µg under the Long-term Conditions**



(b) (4)



(b) (4) Thus, the stability analysis does not support a shelf life of 36 months for Packaging bottle by the worst batch under a common-slope-different-intercept model.

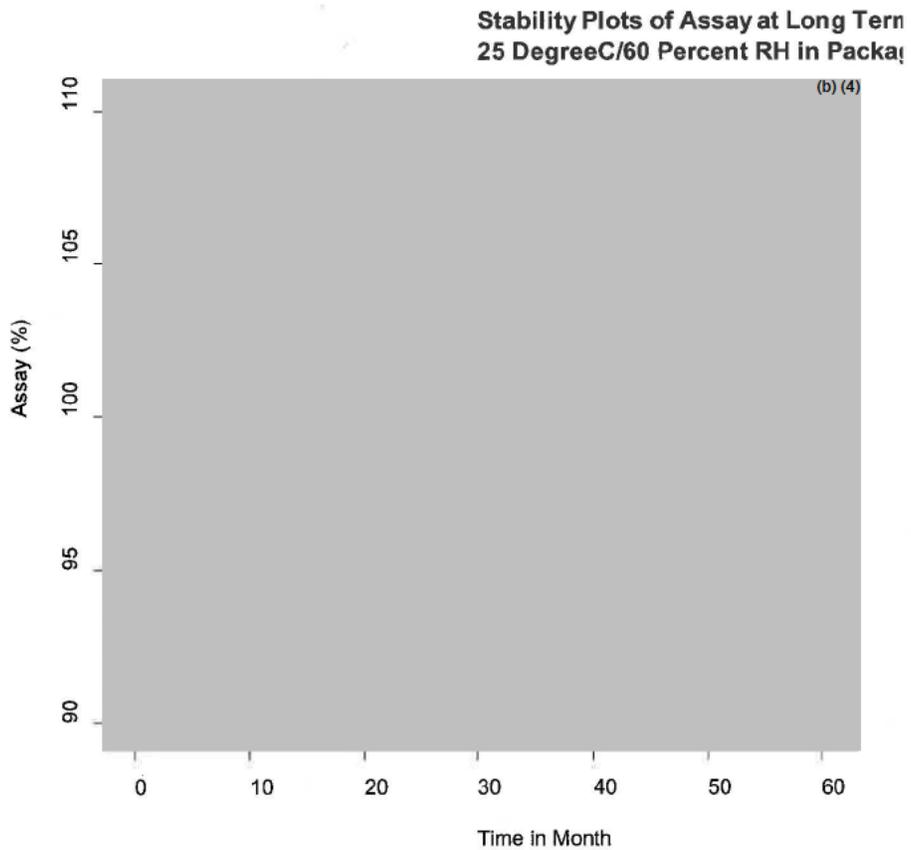
**V.5 Shelf life estimation for dose strength 1600µg**

[Redacted]

**Table 9: Poolability Testing Results for Stability Data of Assay for Packaging Bottle under the Long-term Storage Conditions for Dose Strength 1600µg**

[Redacted Table]

**Figure 5: Stability Plot of Assay for Dose Strength 1600µg under the Long-term Conditions**



[Redacted]

(b) (4) Thus, the stability analysis supports a shelf life of 36 months for Packaging bottle using the pooled data.

## VI. CONCLUSIONS AND RECOMMENDATIONS

The ANCOVA model indicates that the shelf life cannot be determined by the data from certain factors or factor combinations should not be combined. We performed by-batch or pooled analysis for each dose strength level. The analysis results are summarized in the following table.

**Table 10: FDA Statistics Reviewer's Estimated Shelf Life for each Dose Strength using Long-term Stability Data**

Dose strength	Package	Worst Batch	Shelf life estimation in month
200µg	Bottle	12-01963	(b) (4)
400µg	Bottle	12-01965	(b) (4)
800µg	Bottle	12-01968	(b) (4)
1200µg	Bottle	12-01971	(b) (4)
1600µg	Bottle	Pooled	(b) (4)

The 36 months shelf life of Selexipad is supported by the results on dose strength 200µg, 400µg, 800µg and 1600µg if there are no significant changes under the accelerated conditions based on ICH Q1E. (b) (4)



DEPARTMENT OF HEALTH AND HUMAN SERVICES  
PUBLIC HEALTH SERVICE  
FOOD AND DRUG ADMINISTRATION  
CENTER FOR DRUG EVALUATION AND RESEARCH

# STATISTICAL REVIEW AND EVALUATION

## CLINICAL STUDIES

**NDA #:** 207947  
**Drug Name:** UPTRAVI  
**Indication(s):** Pulmonary Arterial Hypertension  
**Applicant:** Actelion Clinical Research, Inc.  
**Date(s):** 12/18/2014  
**Review Priority:** Priority  
**Biometrics Division:** Division of Biometrics I  
**Statistical Reviewer:** Steve Bai, Ph.D.  
**Concurring Reviewers:** James Hung, Ph.D., Director, Division of Biometrics I  
**Medical Division:** Division of Cardiovascular and Renal Products  
**Clinical Team:** Maryann Gordon, MD  
**Project Manager:** Wayne Amchin

# Table of Contents

<b>LIST OF TABLES.....</b>	<b>3</b>
<b>LIST OF FIGURES.....</b>	<b>3</b>
<b>1 EXECUTIVE SUMMARY.....</b>	<b>4</b>
<b>2 INTRODUCTION.....</b>	<b>4</b>
2.1 OVERVIEW .....	4
2.2 DATA SOURCES.....	5
<b>3 STATISTICAL EVALUATION.....</b>	<b>5</b>
3.1 DATA AND ANALYSIS QUALITY .....	5
3.2 EVALUATION OF EFFICACY .....	5
3.2.1 <i>Study Objectives</i> .....	5
3.2.2 <i>Study Design and Endpoints</i> .....	5
3.2.3 <i>Statistical Methodologies</i> .....	7
3.2.4 <i>Patient Disposition, Demographic and Baseline Characteristics</i> .....	11
3.2.5 <i>Results and Exploratory Analyses</i> .....	13
3.2.5.1 Primary Efficacy Results.....	13
3.2.5.2 Secondary Efficacy Analysis.....	18
3.3 EVALUATION OF SAFETY .....	20
<b>4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS.....</b>	<b>20</b>
4.1 AGE, GENDER, AND RACE GROUP .....	20
4.1.1 <i>Age</i> .....	20
4.1.2 <i>Gender</i> .....	21
4.1.3 <i>Race</i> .....	21
4.2 OTHER SUBGROUP POPULATIONS .....	21
4.3 SUBGROUP ANALYSES ON CHANGE FROM BASELINE TO WEEK 26 IN 6MWD .....	22
<b>5 SUMMARY AND CONCLUSIONS.....</b>	<b>23</b>
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE.....	23
5.2 CONCLUSIONS AND RECOMMENDATIONS.....	24

## List of Tables

Table 3-1	Summary of Study Protocol Amendments .....	7
Table 3-2	Summary of group-sequential design .....	8
Table 3-3	Demographic characteristics at screening, FAS .....	11
Table 3-4	Baseline Disease Characteristics, FAS .....	12
Table 3-5	Summary of first CEC-confirmed MM event and components up to 7 days after last study drug intake, FAS .....	13
Table 3-6	Time to Disease Progression in AC-065A302 Treatment Period, by scenario.....	15
Table 3-7	First CEC-confirmed MM event up to Study Closure analysis, FAS.....	16
Table 3-8	Results of Interim Analysis for CEC-confirmed MM events .....	16
Table 3-9	Summary of type of first CEC-confirmed MM event up to 7 days after last study drug intake (including confirmed MM events up to 16 August 2011), FAS .....	17
Table 3-10	Time from randomization to first CEC-confirmed MM event up to 7 days after last study drug intake-in the first 670 randomized patients .....	17
Table 3-11	Competing risk analysis: first CEC-confirmed MM event up to 7 days after last study intake, excluded events with onset date up to 16 August 2011 as events, FAS.....	18
Table 3-12	Absolute change from Baseline to Week 26 in 6MWD at trough .....	19
Table 3-13	Absence of worsening from Baseline in NYHA/WHO functional class at Week 26 (excluding patients with baseline FC IV) .....	20

## List of Figures

Figure 3-1	Kaplan-Meier estimates of time from randomization to first CEC-confirmed MM event up to 7 days after last study drug intake, FAS .....	14
Figure 3-2	Kaplan-Meier estimates of time to death up to Study closure, FAS.....	15
Figure 3-3	Cumulative incidence function for the first CEC-confirmed MM event up to 7 days after last study drug intake.....	18
Figure 4-1	Hazard Ratio (99% CI) for MM events by Age group .....	20
Figure 4-2	Hazard Ratio (99% CI) for MM events by Gender.....	21
Figure 4-3	Hazard Ratio (99% CI) for MM events by Race .....	21
Figure 4-4	Hazard Ratio (99% CI) for MM events by PAH therapy at baseline .....	22
Figure 4-5	Hazard Ratio (99% CI) for MM events by WHO Functional Class at baseline ...	22
Figure 4-6	Hazard Ratio (99% CI) for MM events by Geographical Regions.....	22
Figure 4-7	Subgroup Analyses of Change from Baseline to Week 26 in 6MWD - Part 1.....	23
Figure 4-8	Subgroup Analyses of Change from Baseline to Week 26 in 6MWD - Part 2.....	23

## 1 EXECUTIVE SUMMARY

The efficacy of selexipag in the long-term treatment of Pulmonary Arterial Hypertension (PAH) is based on the data from the pivotal Phase 3 study AC-065A302 (GRIPHON). GRIPHON is the largest clinical outcome study in PAH conducted to date. It was adequately designed as a Morbidity/Mortality (MM) event-driven clinical outcome trial with strict statistical specifications.

The primary objective of the study was met. A statistically highly significant 39% risk-reduction for the occurrence of a first MM event up to end of treatment (EOT) + 7 days was demonstrated with selexipag treatment. The MM event was defined by the first event of death (all causes), hospitalization for PAH worsening, lung transplantation, atrial septostomy, initiation of parenteral prostanoids or chronic oxygen therapy, or disease progression, all adjudicated by an independent CEC blinded to treatment allocation. This represents a clinically highly relevant treatment-effect in a progressive and ultimately fatal cardiovascular disease.

This review does not consider the statistically significant difference between selexipag and placebo in the 6-minute walk distance, secondary symptomatic variable, has any clinical relevance.

## 2 INTRODUCTION

### 2.1 Overview

Selexipag is a selective, orally bioavailable, non-prostanoid agonist at the prostacyclin (PGI<sub>2</sub>) receptor (IP receptor), developed for the treatment of PAH. PAH is characterized by vasculopathy with extensive remodeling of the pulmonary circulation that results in narrowing of the arterial lumen and impaired flow-mediated vasodilation. The consequent increase in pulmonary arterial pressure (PAP) and pulmonary vascular resistance (PVR) limits the ability of the right ventricle to pump blood through the lungs, causing shortness of breath and reduced physical performance. Recent data indicate an average survival of 4 to 5 years after diagnosis in PAH patients with current general medical care and the pharmacological treatment options. There is no cure and PAH remains a progressive and ultimately fatal disorder. Most medicines approved for the treatment of PAH were approved based on their symptomatic effects, evaluated mainly as improvement in exercise capacity in relatively short-term, placebo-controlled studies in selected populations. The only approved PAH-specific medicine with demonstrated benefit for long-term clinical outcome (morbidity/mortality) is the ERA macitentan.

The source of efficacy of this clinical program is based primarily on the results of a single Phase 3 study in the targeted indication. The indication is that selexipag is proposed for the treatment of pulmonary arterial hypertension (PAH). AC-065A302/GRIPHON is the pivotal placebo-controlled, event-driven, group-sequential study, which randomized 1156 patients in a 1:1 ratio to selexipag or placebo. The primary objective of the study was to demonstrate the effect of selexipag on time to first MM event during treatment in patients with PAH.

GRIPHON demonstrated a statistically highly significant 39% risk-reduction for the occurrence of a first MM event up to EOT + 7 days. The observed hazard ratio (versus placebo) was 0.61 with 99% confidence interval of (0.46, 0.81), and 1-sided log-rank p-value < 0.0001.

In addition to the significant finding in the primary endpoint, the difference in median absolute change from Baseline to Week 26 in 6MWD measured at trough was statistically significant between selexipag and placebo (1-sided Wilcoxon-Mann-Whitney  $p = 0.0027$ ). The treatment effect (location shift using Hodges-Lehmann method) versus placebo in the selexipag group was 12.0 m (99% CI: 1, 24).

## 2.2 Data Sources

The sponsor's submitted data are stored in the following directory of the CDER's electronic document room: <\\CDESUB1\evsprod\NDA207947\0000\m5\datasets>.

## 3 STATISTICAL EVALUATION

### 3.1 Data and Analysis Quality

GRIPHON was performed in compliance with GCP guidelines, including the archiving of essential documents. The overall procedures for quality assurance of clinical study data are described in the Actelion Standard Operating Procedures (SOPs). All investigators were trained to comply with GCP and to conduct both studies in accordance with their study protocols.

The reviewer was able to reproduce the results of the primary and key secondary analyses. The applicant submitted the tabulation datasets used to derive the primary analysis dataset. The reviewer was able to trace how the main analysis datasets for the primary and secondary efficacy analyses were derived.

### 3.2 Evaluation of Efficacy

#### 3.2.1 STUDY OBJECTIVES

The primary objective of this study was to demonstrate the effect of selexipag on time to first MM event in patients with PAH.

Secondary objectives include the following:

- To evaluate the effects of selexipag on exercise capacity and other secondary and exploratory efficacy endpoints in patients with PAH.
- To evaluate the safety and tolerability of selexipag in patients with PAH.

#### 3.2.2 STUDY DESIGN AND ENDPOINTS

GRIPHON was a multicenter, randomized, double-blind, parallel group, placebo-controlled, event-driven Phase 3 study to compare the effects of selexipag versus placebo in patients with symptomatic PAH.

A total of 1156 patients were randomized (between 30 December 2009 and 17 May 2013) 1:1 to selexipag or placebo, with stratification by site and a block size of 4 at 181 sites in 39 countries.

The study included the following trial periods:

- Screening period: performed up to a maximum 28 days before baseline visit (Visit 1).
- Treatment period: Visit 1 to end of study (EOS) visit. This period was concluded with an EOS visit at the time of study closure announcement (i.e. once the overall target number of 331 CEC-confirmed MM events with onset date up to 7 days after last study drug was achieved)
- Post-treatment observation period (PTOP): patients who discontinued study drug with or without an MM event prior to Study closure announcement had an option to enter a post-treatment observation period to collect additional clinical data. This was defined as the period after the EOS visit following discontinuation of study drug up to the post-treatment observation closure visit (PTOCV) following the announcement of Study closure by Actelion.

### Study Population

Patients with PAH in modified NYHA/WHO FC I–IV were included in this study. NYHA/WHO FC I and II patients were included in order to investigate the occurrence of clinical events in a population with less advanced disease. Adults ( $\geq 18$  years) including elderly patients (up to 75 years inclusive) at time of entry into GRIPHON with symptomatic PAH, either naïve to or receiving PAH specific treatment (ERAs and/or PDE-5i), were included.

### Treatment

The treatment groups are assigned with either filmed-coated tablets containing 200 $\mu$ g selexipag or matching placebo. The eligible patients were randomized in a 1:1 ratio to selexipag or placebo using a centralized randomization system via IWRS/IVRS.

### Efficacy Endpoints

The primary efficacy endpoint was time to first CEC-confirmed MM event up to 7 days after the last study drug intake in the treatment period (i.e., end of treatment [EOT] + 7 days). The following MM events were considered:

- Death (all-causes),
- Hospitalization for worsening of PAH based on predefined criteria
- Worsening of PAH resulting in need for lung transplantation or balloon atrial septostomy
- Initiation of parenteral prostanoid therapy or chronic oxygen therapy due to worsening of PAH
- Disease progression (patients in modified NYHA/WHO FC II/III at baseline) confirmed by
  - Decrease in 6MWD from Baseline ( $\geq 15\%$ , confirmed by 2 tests on different days) **and**
  - Worsening of NYHA/WHO FC
- Disease progression (patients in NYHA/WHO FC III/IV at baseline) confirmed by:
  - Decrease in 6MWD from Baseline ( $\geq 15\%$ , confirmed by 2 tests on different days) **and**
  - Need for additional PAH-specific therapy.

Patients in NYHA/WHO FC III at baseline qualified for both disease progression definitions.

There are following secondary endpoints:

- Absolute change from Baseline to Week 26 in 6MWD measured at trough. Prior to implementation of Amendment 1, this was the primary endpoint.
- Absence of worsening from Baseline to Week 26 in NYHA/WHO FC.
- Time from randomization to first of CEC-confirmed death due to PAH or CEC-confirmed hospitalization due to PAH worsening up to 7 days after last study drug intake.
- Time from randomization to death of all causes up to Study closure.
- Absolute change from Baseline to Week 26 in the sub-scale ‘Breathlessness’ of CAMPHOR (Cambridge Pulmonary Hypertension Outcome Review) ‘Symptoms’ (at selected centers).
  - The sub-scale ‘Breathlessness’ of CAMPHOR ‘Symptoms’ was defined as the sum of the ‘Breathlessness’ items 11 to 18. It ranged from 0 (good) to 8 (poor).
- Absolute change from Baseline to Week 26 in CAMPHOR ‘Symptoms’ score (at selected centers).
  - The CAMPHOR ‘Symptoms’ score was defined as the sum of the ‘Symptoms’ items 1 to 25. It ranged from 0 (good) to 25 (poor).

### 3.2.3 STATISTICAL METHODOLOGIES

The original study protocol was written on September 17, 2009. The primary endpoint in this study was “Change from Baseline to Week 16 in 6MWD.” Actelion submitted study’s first global protocol amendment to FDA on March 11, 2010. The final version, Amendment 6, was submitted on January 23, 2013. Table 3-1 listed main statistical changes of each protocol amendments.

**Table 3-1 Summary of Study Protocol Amendments**

Versions	Date	Key Statistical Changes
Amendment 1	3/11/2010	1. Changed the primary endpoint to <i>time to first clinical worsening</i> . 2. Moved <i>change from baseline to week 16 in 6MWD</i> to secondary endpoint and changed to be assessed at Week 26. 3. Merged study AC-065A301 into AC-06A302, and renamed it as AC-065A302/GRIPHON
Amendment 2	12/20/2010	Fixed Type II error typo from 0.01 to 0.1, i.e. power=90%
Amendment 3	05/11/2011	No major statistical related changes
Amendment 4	08/10/2011	1. Changed initial target hazard ratio of 0.5729 to 0.65 2. The amended treatment effect increased the number of primary events from 202 to 332. The sample size was increased from 670 to 1150 patients. 3. Added an interim analysis after observing 202 primary events.
Amendment 5	12/14/2011	1. In order to eliminate any concern that the protocol change (Amendment 4) could be considered informed, the 46 events observed until 16 August 2011 are censored in the primary endpoint analysis. 2. For the interim analysis, the type-I error was set to 0.0001 (1-sided) according to the Haybittle-Peto stopping rule. 3. The main sensitivity analysis of the primary endpoint to be performed on 378 events (including the censored events up to 16 August 2011).
Amendment 6	01/23/2013	The ‘ <i>clinical worsening event</i> ’ was replaced by ‘ <i>morbidity and mortality (MM) event</i> ’

[Source: Sponsor’s Protocol and amendments documents]

There are a total of five versions of Statistical Analysis Plans. The SAP versions 1.0 and 2.0 were both written in April of 2013 and prior to the proposed interim analysis. The cut-off date for interim analysis on the confirmation of MM events by CEC was March 25, 2013. The SAP versions 3.0, 4.0 and 5.0 were issued in February and June of 2014 and were all prior to unblinding of the final analysis. The final analysis described in this report covers all data from study AC-065A302 up to the cut-off date of 27 April 2014, which is the date of last patient visit.

### Sample Size Assumptions

In order to determine sample size and power, the following assumptions have been made:

- Constant hazard rate of 0.2231 per year for placebo
- Constant hazard rate of 0.145 per year for selexipag (hazard ratio of 0.65)
- Constant hazard rate of 0.051 per year for drop-out (end of study without a Morbidity/mortality event) in both treatment groups

These lead to a total of 331 morbidity/mortality events confirmed by the CEC in order to obtain an overall power of 90% for rejection of the null hypothesis. These assumptions were consistent throughout all five versions of SAP.

### Efficacy Analysis Methods

The Full analysis Set (FAS) included all randomized patients in GRIPHON. All main statistical analyses of all efficacy endpoints were based on the FAS. Study GRIPHON employed a group-sequential design for the primary efficacy endpoint with options to recommend stop for futility or for compelling and robust efficacy at the interim analysis.

The design was pre-specified since SAP version 1.0 and did not have any major statistical modifications in the subsequent revisions. The design features are:

- the one-sided overall type-I-error probability is fixed to  $\alpha = 0.005$
- the maximum information is specified as 331 MM events
- the one-sided type-I-error probability at the interim analysis is fixed to 0.00005
- the interim analysis is planned with 202 MM events (an information fraction of  $202/331 = 61\%$ ); however, if there are less or more events available for the interim analysis, the one-sided significance level used for the interim analysis is kept unchanged at 0.00005.

The one-sided log-rank statistic  $Z$  was used to test the null hypothesis for time to first MM event (See next subsection for detail). In addition, a non-binding interim futility stopping rule was foreseen as outlined in Table 3-2 below.

**Table 3-2 Summary of group-sequential design**

Analysis stage (anticipated cumulative number of events <sup>1</sup> )	Efficacy			Futility		
	Cumulative alpha <sup>2</sup> spent	Guidance <sup>3</sup> to reject H <sub>0</sub>		Cumulative beta spent	Guidance <sup>3</sup> to accept H <sub>0</sub>	
		<i>p</i> -value	Z-score		<i>p</i> -value	Z-score
Interim (202 events)	0.00005	≤ 0.00005	≥ 3.8906	0.0013	≥ 0.5	≤ 0
Final (331 events)	0.005	≤ 0.004991	≥ 2.5764	0.1	> 0.004991	< 2.5764

At the interim and final analysis stage, the following statistics were calculated:

- Z statistic for the log-rank test.
- Point estimate and standard error of the log hazard ratio from the proportional hazard model.

If at the interim stage the observed Z-score was greater or equal to the efficacy stopping boundary for rejection of the null hypothesis (if Z-score  $\geq 3.8906$ ), the DMC could nevertheless recommend to continue the trial to its end, i.e., when 331 morbidity/mortality events had been confirmed by the CEC.

### Analysis of primary efficacy endpoint

Hypotheses were formulated in terms of “survival” functions  $S(t)$ , i.e., the probability that time to first MM event was  $\geq t$  for a day  $t$ . One-sided hypotheses were expressed in terms of the survival functions  $S_{m/m,selexipag}(t)$  and  $S_{m/m,placebo}(t)$  for selexipag and placebo, respectively.

$H_{0,m/m}$ :  $S_{m/m,selexipag}(t) \leq S_{m/m,placebo}(t)$  for all  $t \geq 0$  versus

$H_{A,m/m}$ :  $S_{m/m,selexipag}(t) \geq S_{m/m,placebo}(t)$  for all  $t \geq 0$  and  $S_{m/m,selexipag}(t) > S_{m/m,placebo}(t)$  for some  $t > 0$ .

The primary analysis was performed on the FAS using a one-sided unstratified log-rank test. The unstratified log-rank test is conducted by SAS PROC LIFETEST where the STRATA statement includes only the treatment group variable.

No imputation method was applied. For a patient without a MM event up to 7 days after last study drug intake in the treatment period, time to first MM event was defined using the following censoring rules:

- For randomized patients who received at least one intake of study drug and who did not consent to the post-treatment observation period: minimum (date of last study drug intake plus 7, EOS visit date, date of last contact, analysis cut-off date, i.e., 27 April 2014) minus date of randomization plus 1.
- For randomized patients who received at least one intake of study drug and who did consent to the post-treatment observation period: minimum (date of last study drug intake plus 7, date of last contact, 27 April 2014) minus date of randomization plus 1.
- For randomized patients who did not receive any study drug: minimum (maximum [EOS visit date, randomization date], date of last contact, 27 April 2014) minus date of randomization plus 1.

Following Protocol Amendment 5, CEC-confirmed MM events with onset date up to 16 August 2011 were considered as censored at the event onset date for the primary statistical analysis. In the event that a patient with a CEC-confirmed MM event with onset date up to 16 August 2011 had a subsequent CEC-confirmed MM event with onset date after 16 August 2011, then the first event was disregarded and the second event was counted as an event in the statistical analysis.

As supportive evidence, competing risks analysis was described for the primary efficacy endpoint in statistical analysis plans. For each patient, a multistate stochastic process  $(X_t)_{t \geq 0}$  is considered where  $X_t$  denotes the state a patient is at day  $t$ ,  $0 \leq t \leq$  date of last study drug intake plus 7 days. Each patient would fall into one of following 5 states:

[Death]	death (all-cause mortality)
[Hosp]	hospitalization for worsening of PAH based on predefined criteria
[L/B/P/O2]	worsening of PAH resulting in need for lung transplantation or balloon atrial septostomy, parenteral prostanoid treatment or chronic oxygen therapy
[DP]	disease progression
[No event]	no event

The above event components of the primary endpoint will be analyzed by a competing risk methodology as proposed by Gray<sup>1</sup>. Cumulative incidence functions for the time to first event and Gray's test will be calculated using the R-package *cmprsk*.

### **Analysis of secondary efficacy data**

In case of rejection of the null hypothesis in the main statistical analysis of the primary efficacy endpoint, the null hypotheses for the secondary efficacy endpoints were tested in a conditional hierarchical manner. A null hypothesis was rejected if and only if the main analysis of the endpoint and all main analyses of preceding secondary efficacy endpoints resulted in rejection of respective null hypotheses. This procedure restricted the family-wise type-I-error rate to the overall assigned one-sided alpha of 0.005.

#### Absolute change from baseline in 6MWD at trough at Week 26:

The main analysis was performed on the FAS. The following non-parametric ANCOVA procedure was used:

1. Transformation of Baseline and post-Baseline values for all patients (regardless of treatment groups) to standardized ranks (i.e., ranks divided by the number of patients ranked plus 1, mean ranks in case of ties)
2. Determination of residuals from the linear regression of the response variable standardized ranks on Baseline variable standardized ranks.
3. Application of the one-sided Wilcoxon-Mann-Whitney test to these residuals. The standardized test statistic with a continuity correction of 0.5 is asymptotically standard normally distributed under the null hypothesis. One-sided significance level of 0.005 is used.

For patients without any 6MWD available at Week 26, the following imputation algorithm was applied:

Rule 1: for patients unable to walk at Week 26, this included the following:

- patients who died before study Day 271 (upper limit of the Week 26 time window) without any visit performed in the Week 26 time window,
- patients for whom the Week 26 visit corresponded to a clinical worsening event (CWE) visit and who were unable to walk for PAH reasons, i.e., reason was 'Dyspnea/Fatigue' or reason was coded as 'Related to Pulmonary Arterial Hypertension'),

a value of 0 was imputed for 6MWD at Week 26.

Rule 2 (if rule 1 did not apply): the second lowest observed 6MWD value at Week 26 in the same analysis set, irrespective of study treatment group, was imputed. In the FAS, this was 10 m.

---

<sup>1</sup> Gray RJ (1988). A class of k-Sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* 16: 1141–54.

This missing data imputation was handled with multiple imputations before SAP version 4.0. The current single imputation approach was the advice that sponsor received from FDA.

*Absence of worsening from baseline in NYHA/WHO FC at Week 26:*

The main analysis was performed on the FAS excluding patients in FC IV at Baseline. A Cochran-Mantel-Haenszel test stratified by NYHA/WHO FC at Baseline was used. For patients with missing NYHA/WHO FC at Week 26, the NYHA/WHO FC was considered as having worsened from Baseline at Week 26 in the main analysis.

*Other Secondary endpoints:* The main analyses for the remaining time to event and continuous secondary endpoints were tested similarly as the primary endpoint and key secondary endpoint, respectively. In the FAS, the statistical significance stopped at *Absence of worsening from baseline in NYHA/WHO FC at Week 26*. This review did not present the results of any further secondary endpoints.

### 3.2.4 PATIENT DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS

A total of 1351 patients at 181 sites in 39 countries were screened in study GRIPHON. These patients were randomized in a 1:1 ratio to selexipag (N=574) or placebo (N=582). 195 (14.4%) patients failed screening and did not receive study treatments. The most frequently reported reasons for screening failure were 6MWD not within the allowed protocol-specified range (44 subjects) and moderate or severe obstructive lung disease (19 subjects).

**Table 3-3 Demographic characteristics at screening, FAS**

	Selexipag N=574	Placebo N=582	Total N=1156
Sex			
Male (%)	117 (20.4)	116 (19.9)	233 (20.2)
Female (%)	457 (79.6)	466 (80.1)	923 (79.8)
Age			
Mean (SD)	48.2 (15.2)	47.9 (15.6)	48.1 (15.4)
< 65 (%)	475 (82.8)	474 (81.4)	949 (82.1)
65-74 (%)	91 (15.9)	103 (17.7)	194 (16.8)
≥75 (%)	8 (1.4)	5 (0.9)	13 (1.1)
BMI			
Mean (SD)	26.9 (6.4)	26.7 (6.1)	26.8 (6.3)
Race (%)			
White	376 (65.5)	375 (64.4)	751 (65.0)
Asian	125 (21.8)	120 (20.6)	245 (21.2)
Black	13 (2.3)	14 (2.4)	27 (2.3)
Hispanic	51 (8.9)	63 (10.8)	114 (9.9)
Other	9 (1.6)	10 (1.7)	19 (1.6)

[Source: Reviewer's Results]

Of the 1156 randomized patients, 1152 patients (selexipag: 574 [100%], placebo: 578 [99.3%]) received study treatment during the treatment period. The reasons that four placebo patients did not receive study treatment are listed as following:

- 1 placebo patient had a CEC-confirmed MM event 1 day after randomization to the placebo group and did not receive study treatment.
- 2 placebo patients randomized to placebo did not receive study treatment due to withdrawal of consent from all study components;
- 1 placebo patient randomized to placebo did not receive study treatment due to an administrative reason.

The demographic characteristics in the FAS were comparable between selexipag and placebo groups [Table 3-3]. The study population predominantly comprised females (79.8%). Average age at screening was 48.1 years, with 82.1% of patients < 65 years old. Mean BMI was 26.8 kg/m<sup>2</sup>. The majority of patients were Caucasian/white (65.0%) or Asian (21.2%).

**Table 3-4 Baseline Disease Characteristics, FAS**

	Selexipag N=574	Placebo N=582	Total N=1156
Time since PAH (years) Mean (SD)	2.3 (3.5)	2.5 (3.8)	2.4 (3.6)
NYHA/WHO Functional class (%)			
I	4 (0.7)	5 (0.9)	9 (0.8)
II	274 (47.7)	255 (43.8)	529 (45.8)
III	293 (51.0)	314 (54.0)	607 (52.5)
IV	3 (0.5)	8 (1.4)	11 (1.0)
6 MWD (m) Mean (SD)	358.5 (76.3)	348.0 (83.2)	353.2 (80.0)
Borg Dispnea Index Mean (SD)	3.6 (2.1)	3.7 (2.1)	3.7 (2.1)
SBP (mmHg) Mean (SD)	115.0 (16.2)	114.1 (15.4)	114.5 (15.8)
DBP (mmHg) Mean (SD)	72.3 (10.7)	71.9 (10.4)	72.1 (10.5)
Heart Rate (beats/min) Mean (SD)	77.3 (12.3)	77.1 (11.8)	77.2 (12.1)

[Source: Reviewer's Results]

Baseline disease characteristics in the FAS were generally comparable across the selexipag and placebo groups [Table 3-4]. On average, patients' time since PAD diagnosis was 2.4 years. At baseline, patients were predominantly in NYHA/WHO FC II (45.8%) and FC III (52.5%). Mean 6MWD at baseline was 353.2 m (range: 50–515 m). Mean Borg dyspnea index was 3.7 (range: 0–10.0).

### 3.2.5 RESULTS AND EXPLORATORY ANALYSES

#### 3.2.5.1 Primary Efficacy Results

There were major changes to the protocol based on global Amendment 4 as described in Table 3-1, which was submitted to the FDA on August 16, 2011. In order to eliminate any concern that these changes could be considered informative, any MM events with a CEC-confirmed onset date up to 16 August 2011 were censored and were not considered in the main analysis of the primary endpoint. However, the results for the primary endpoint with and without censoring of CEC-confirmed MM events up to 16 August 2011 were very similar.

#### Main analysis results

In the main analysis, 47 CEC-confirmed MM events in 47 patients (16 in selexipag and 31 in placebo) with an onset date up to 16 August 2011 were excluded. Subsequently, 2 of the 47 patients (Placebo Patient 1901-21321 and Selexipag Patient 6006-20241) had a CEC-confirmed MM event after 16 August 2011. Therefore 45 patients were actually censored for the main analysis due to occurrence of an MM event up to 16 August 2011.

**Table 3-5 Summary of first CEC-confirmed MM event and components up to 7 days after last study drug intake, FAS**

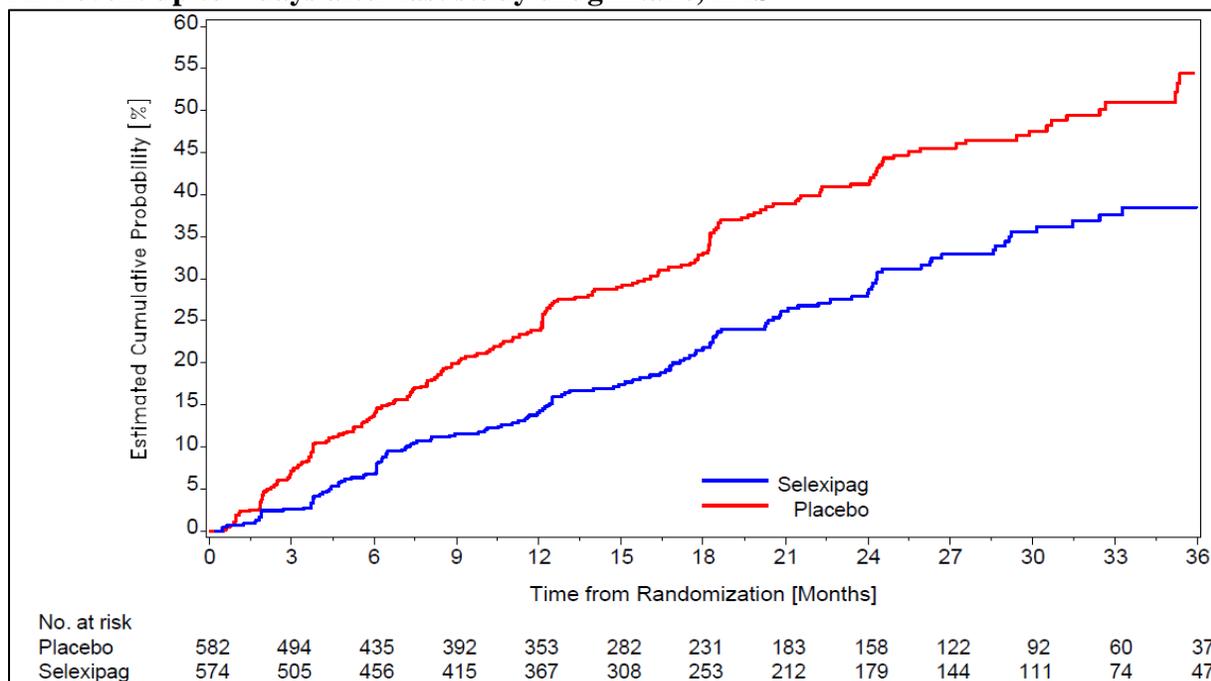
Event	Selexipag N=574		Placebo N=582		Hazard Ratio (99% CI)	p-value
	n	%	n	%		
<b>Final Analysis</b>					0.61	
First morbidity/mortality event	140	24.4	212	36.4	[0.46,0.81]	<0.0001
<b>Decomposition of the first MM event</b>						
• <b>Death</b>	25	4.4	16	2.7		
• <b>Hospitalization</b> for PAH worsening	71	12.4	96	16.5		
• <b>PAH</b> worsening resulting in need for lung transplantation or balloon atrial septostomy	1	0.2	2	0.3		
• <b>Parenteral</b> prostanoid therapy or chronic oxygen therapy	11	1.9	14	2.4		
• <b>Disease</b> progression	32	5.6	84	14.4		
<b>First Occurrence of each component of MM event</b>						
• <b>Death</b>	40	7.0	34	5.8	1.10 [0.61,2.01]	
• <b>Hospitalization</b> for PAH worsening	77	13.4	111	19.1	0.65 [0.44,0.95]	
• <b>PAH</b> worsening resulting in need for lung transplantation or balloon atrial septostomy	2	0.3	3	0.5	0.63 [0.06,6.67]	
• <b>Parenteral</b> prostanoid therapy or chronic oxygen therapy	30	5.2	45	7.7	0.62 [0.34, 1.14]	
• <b>Disease</b> progression	58	10.1	127	21.8	0.43 [0.29,0.65]	

[Source: Reviewer's Results]

The CEC-confirmed MM event up to 7 days after last study drug intake in the treatment period was recorded for 140 patients in the selexipag group compared to 212 patients in the placebo group. In the final main time-to-event analysis the 1-sided unstratified log-rank p-value was < 0.0001. The hazard ratio for selexipag versus placebo for the occurrence of an MM event was 0.61 with a 99% confidence interval of (0.46, 0.81). The corresponding relative risk reduction with selexipag versus placebo was 39% [Table 3-5].

The KM estimates for experiencing an MM event were consistently lower in the selexipag group than in the placebo group throughout the treatment period [Figure 3-1].

**Figure 3-1 Kaplan-Meier estimates of time from randomization to first CEC-confirmed MM event up to 7 days after last study drug intake, FAS**



[Source: Reviewer’s Results]

The proportion of patients who experienced a CEC-confirmed MM event up to 7 days after last study drug intake was 24.4% in the selexipag group compared to 36.4% in the placebo group. Hospitalization for PAH worsening was the most frequently reported first CEC-confirmed MM event, occurring in 12.4% of patients in the selexipag group and 16.5% in the placebo group.

Disease progression as the first event was recorded for 5.6% of patients in the selexipag group compared to 14.4% in the placebo group. Death as the first event was recorded for 4.4% of patients in the selexipag group and 2.7% in the placebo group.

The Disease progression defined as the following two scenarios:

- a. Decrease in 6MWD from Baseline ( $\geq 15\%$ , confirmed by 2 tests on different days within 2 weeks) and Worsening of NYHA/WHO FC

- b. Decrease in 6MWD from Baseline ( $\geq 15\%$ , confirmed by 2 tests on different days within 2 weeks) and Need for additional PAH-specific therapy.

Disease progression is the most influential component of the MM event. Table 3-6 examined the effects of above two scenarios in the Disease progression. Patients in the selexipag group showed much lower risks in both scenarios than patients in the placebo group. Furthermore, the treatment effects are slightly larger under scenario a. There are number of disease progressions classified in both scenarios.

**Table 3-6 Time to Disease Progression in AC-065A302 Treatment Period, by scenario**

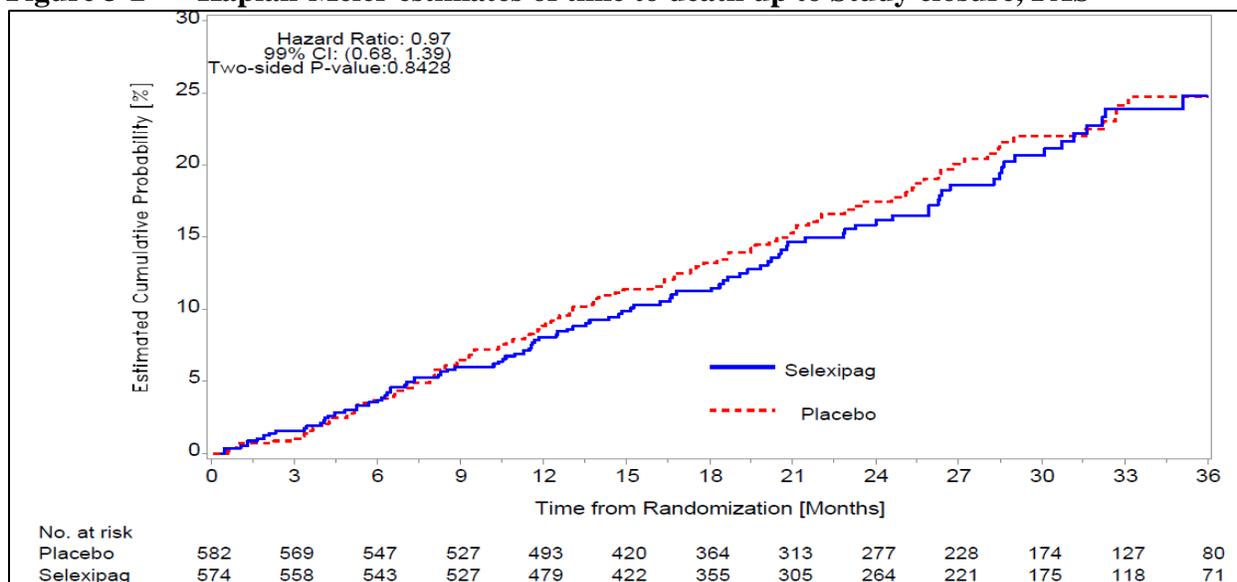
Disease Progression	Selexipag	Placebo	HR
First Occurrence	58	127	0.43
Scenario a	37	88	0.40
Scenario b	44	88	0.47
Scenario a and b	23	49	
Decomposition of MM event	32	84	0.36
Scenario a	17	54	0.30
Scenario b	28	56	0.47
Scenario a and b	13	26	

[Source: Sponsor’s response and confirmed by reviewer]

**Time to MM events and components up to Study Closure (ITT) Analyses**

In the main analysis of the primary endpoint, selexipag group observed notably higher rates of Death and Death as the first MM event than placebo group, see Table 3-5. However, the concern of adverse effect in morality is alleviated by the time to all-cause death up to study closure analysis. A total of 100 and 105 patients in the selexipag and placebo groups, respectively, died up to study closure. The hazard ratio for selexipag versus placebo for the time to death up to study closure was 0.97 (99% CI: 0.68, 1.39), see Figure 3-2.

**Figure 3-2 Kaplan-Meier estimates of time to death up to Study closure, FAS**



[Source: Reviewer’s result]

In the Intent-to-Treat analysis, a total of 185 patients in the selexipag group compared to 258 patients in the placebo group had a CEC-confirmed MM event up to study closure [Table 3-7]. The hazard ratio for selexipag versus placebo for the occurrence of an MM event up to study closure in the FAS was 0.65 (99% CI: 0.54, 0.79). These results are consistent with the main analysis of primary analysis.

**Table 3-7 First CEC-confirmed MM event up to Study Closure analysis, FAS**

Event	Selexipag N=574		Placebo N=582		Hazard Ratio (99% CI)	p-value
	n	%	n	%		
<b>Final Analysis</b>						
First morbidity/mortality event	185	32.2	258	44.3	0.65 [0.54,0.79]	<0.0001
<b>Decomposition of the first MM event</b>						
• <b>Death</b>	45	7.8	28	4.8		
• <b>Hospitalization</b> for PAH worsening	86	15.0	114	19.6		
• <b>PAH</b> need lung transplantation	1	0.2	2	0.3		
• <b>Parenteral</b> prostanoid therapy	15	2.6	14	2.4		
• <b>Disease</b> progression	38	6.6	100	17.2		
<b>First Occurrence of each component of MM event</b>						
• <b>Death</b>	100	17.4	105	18.0	0.97 [0.68, 1.39]	
• <b>Hospitalization</b> for PAH worsening	100	17.4	137	23.5	0.70 [0.50, 0.99]	
• <b>PAH</b> need lung transplantation	5	0.9	11	1.9	0.45 [0.11, 1.80]	
• <b>Parenteral</b> prostanoid therapy	47	8.2	63	10.8	0.73 [0.44, 1.20]	
• <b>Disease</b> progression	67	11.7	147	25.3	0.43 [0.30, 0.64]	

[Source: Sponsor's Response to FDA request and confirmed by reviewer]

### **Results of Interim Analysis**

The planned interim analysis had a data cut-off date of 25 March 2013. The observed one-sided log-rank p-value was 0.00010. Given the pre-specified interim analysis one-sided nominal alpha of 0.00005, the statistical stopping criterion was almost very nearly met [Table 3-8].

**Table 3-8 Results of Interim Analysis for CEC-confirmed MM events**

Event	Selexipag N=548		Placebo N=564		Hazard Ratio (99.99% CI)	p-value
	n	%	n	%		
<b>Interim Analysis</b>						
First morbidity/mortality event	89	16.2	141	25.0	0.60 [0.35,1.01]	0.0001

[Source: Reviewer's Results]

### **Analysis including events onset date up to 16 August 2011**

As a sensitive analysis, the following analysis included events with CEC-confirmed onset date up to 16 August 2011. The CEC-confirmed MM event up to 7 days after last study drug intake was recorded for 155 patients in the selexipag group compared to 242 patients in the placebo group. In the time-to-event analysis, the hazard ratio for selexipag versus placebo for the occurrence of an MM event was 0.60 (99% CI: 0.46, 0.78, 1-sided unstratified log-rank  $p <$

0.0001). The corresponding relative risk reduction with selexipag versus placebo was 40% [Table 3-9].

**Table 3-9 Summary of type of first CEC-confirmed MM event up to 7 days after last study drug intake (including confirmed MM events up to 16 August 2011), FAS**

Event	Selexipag N=574		Placebo N=582		Hazard Ratio (99% CI)	p-value
	n	%	n	%		
First morbidity/mortality event	155	27.0	242	41.6	0.60 [0.46,0.78]	<0.0001
Death	28	4.9	18	3.1		
Hospitalization for PAH worsening	78	13.6	109	18.7		
PAH need lung transplantation	1	0.2	2	0.3		
Parenteral prostanoid therapy	10	1.7	13	2.2		
Disease progression	38	6.6	100	17.2		

[Source: Reviewer's Results]

#### **Analysis first CEC-confirmed MM events in the first 670 randomized patients**

A supportive analysis was performed in the first 670 randomized patients up to the occurrence of 202 first CEC-confirmed MM events, representing the originally planned study design prior to Study Protocol Amendment 5. For this analysis, a CEC-confirmed MM event was recorded for 81 patients in the selexipag group compared to 121 patients in the placebo group. The hazard ratio for selexipag versus placebo for the occurrence of an MM event was 0.65 with a 99% CI of (0.45, 0.94) and p-value is 0.0014 [Table 3-10]. This supportive analysis is consistent with that of the main analysis on the primary endpoint.

**Table 3-10 Time from randomization to first CEC-confirmed MM event up to 7 days after last study drug intake-in the first 670 randomized patients**

Event	Selexipag N=324		Placebo N=346		Hazard Ratio (99% CI)	p-value
	n	%	n	%		
First morbidity/mortality event	81	25.0	121	35.0	0.65 [0.45,0.94]	0.0014

[Source: Reviewer's Results]

#### **Competing risk analysis**

In order to gain better understanding of the treatment effect on each component of the primary endpoint, section 3.2.3 described Gray's competing risk analysis as supportive evidence. All 4 MM event components (Death, DP, Hosp and L/B/P/O2) competed with each other during the treatment period since the occurrence of one prevented the observation of the others up to 7 days after last study drug intake.

Patients in the selexipag group showed a lower risk of disease progression ( $p < 0.0001$ ) than patients in the placebo group. Nearly significant difference was observed for the risk of hospitalization for PAH worsening ( $p=0.0635$ ). No significant difference was observed between selexipag and placebo for the risk of death ( $p = 0.1172$ ) or for the risk of PAH worsening ( $p =$

0.5107). The finding in Table 3-11 are consistent with results of Table 3-5, where the result of MM event was primarily driven by the components of disease progression and hospitalization for PAH worsening.

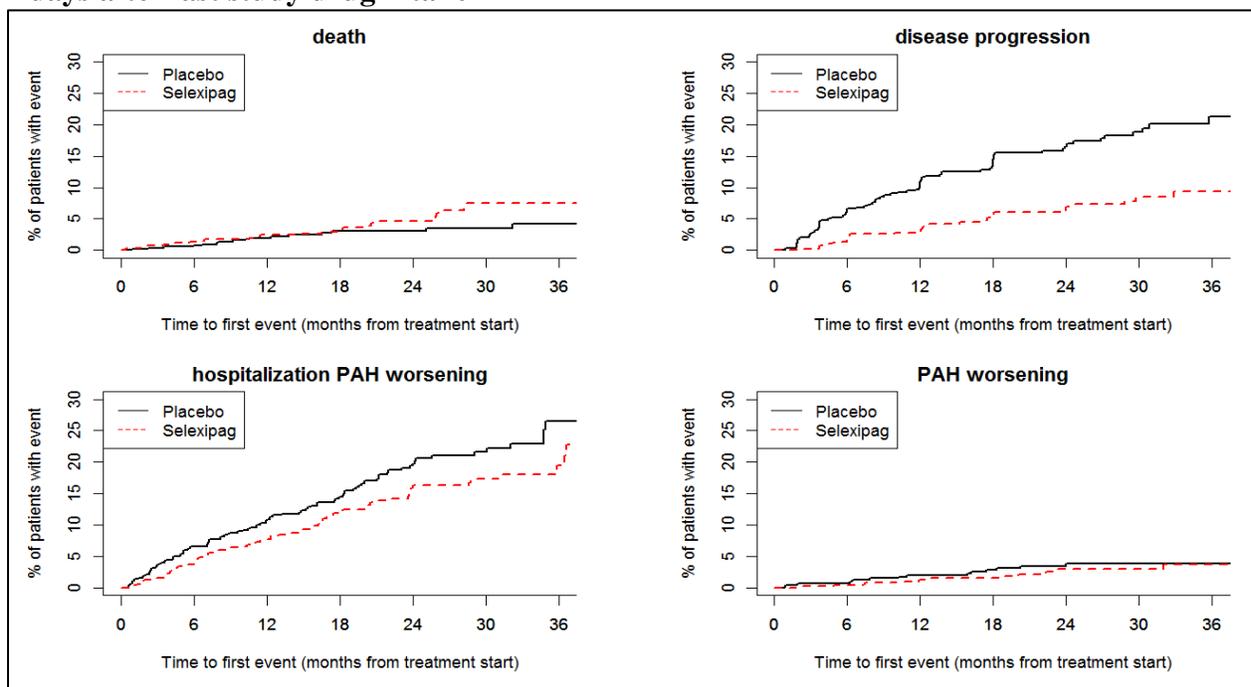
**Table 3-11 Competing risk analysis: first CEC-confirmed MM event up to 7 days after last study intake, excluded events with onset date up to 16 August 2011 as events, FAS**

Treatment differences (Gray's test)		
Selexipag vs Placebo:	Chi-square	P-value
Death	2.4550	0.1172
Hospitalization for PAH worsening	3.4434	0.0635
PAH worsening	0.4326	0.5107
Disease progression	24.1529	0.0000

[Source: Reviewer's Results]

Cumulative incidence functions for each event by treatment group are presented in Figure 3-3.

**Figure 3-3 Cumulative incidence function for the first CEC-confirmed MM event up to 7 days after last study drug intake**



[Source: Reviewer Results]

### 3.2.5.2 Secondary Efficacy Analysis

#### Absolute change from Baseline to Week 26 in 6MWD

Median absolute change from Baseline to Week 26 in 6MWD measured at trough was 4.0 m in the selexipag group and -9.0 m in the placebo group [Table 3-12]. In the main analysis using a non-parametric ANCOVA with covariate 6MWD at Baseline, the difference was

statistically significant (1-sided Wilcoxon-Mann-Whitney  $p = 0.0027$ ). The treatment effect (location shift using Hodges-Lehmann method) versus placebo for selexipag was 12 m.

Even though, 6MWD showed statistically significant difference favoring selexipag over placebo. However, both treatment groups worsened the walk distance at Week 26. Selexipag and placebo group walked 52.0 m and 66.3 m less than baseline, respectively.

**Table 3-12 Absolute change from Baseline to Week 26 in 6MWD at trough**

<b>6 Minutes Walking Test (m)</b>	Selexipag N=574	Placebo N=582
<b>Baseline</b>		
Mean (SD)	358.5 (76.3)	348.0 (83.2)
Median	376.0	369.0
<b>Week 26</b>		
Mean (SD)	306.5 (170.0)	281.7 (173.8)
Median	370.0	346.0
<b>Absolute change from Baseline at Week 26</b>		
Mean (SD)	-52.0 (150.2)	-66.3 (148.2)
Median	4.0	-9.0
<b>Main Analysis (Non-parametric ANCOVA)</b>		
Hodges-Lehmann location shift (99% CI)	12.0 (1, 24)	
Wilcoxon-Mann-Whitney Statistic	2.786	
p-value	0.0027	

[Source: Reviewer's Results]

The main analysis on the change from Baseline to Week 26 in 6MWD measured at trough showed a statistically significant difference favoring selexipag over placebo, so the next secondary endpoint can also be tested.

#### **Absence of worsening from Baseline in NYHA/WHO FC at Week 26**

Absence of worsening from Baseline in NYHA/WHO functional class at Week 26 was reported for 77.8% of patients in the selexipag group and 74.9% in the placebo group [Table 3-13]. For the analysis on the absence of worsening from Baseline in NYHA/WHO FC, patients with FC IV at baseline were excluded as they could not shift to a worse category. In the main analysis using a 2-sided Cochran-Mantel-Haenszel test stratified by NYHA/WHO FC at Baseline, a p-value of 0.2843 was obtained. The Breslow-Day test had a  $p = 0.1916$ , which meant the common odds ratio is a valid measure. The common odds ratio for the effect of selexipag relative to placebo was 1.161 with a 99% CI of (0.811, 1.664).

The main analysis of absence of worsening did not show a statistically significance difference between selexipag and placebo. Therefore, this review follows the pre-specified statistical analysis plan and does not examine the efficacy of any additional secondary endpoints.

**Table 3-13 Absence of worsening from Baseline in NYHA/WHO functional class at Week 26 (excluding patients with baseline FC IV)**

Baseline		Week 26	
Selexipag	N	Absence of Worsening n (%)	Worsening n (%)
I	4	4 (100%)	0 (0%)
II	274	214 (78.1%)	60 (21.9%)
III	293	226 (77.1%)	67 (22.9%)
All	571	444 (77.8%)	127 (22.2%)
Placebo	N		
I	5	4 (80%)	1 (20%)
II	255	204 (80%)	51 (20%)
III	314	222 (70.7%)	92 (29.3%)
All	574	430 (74.9%)	144 (25.1%)
CMH statistic (p-value)		1.147 (0.2843)	
Odds Ratio (99% CI)		1.161 (0.811, 1.664)	
Breslow-Day statistic (p-value)		3.304 (0.1916)	

[Source: Reviewer’s Results]

### 3.3 Evaluation of Safety

Safety is not evaluated in this review. Please see the clinical review.

## 4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

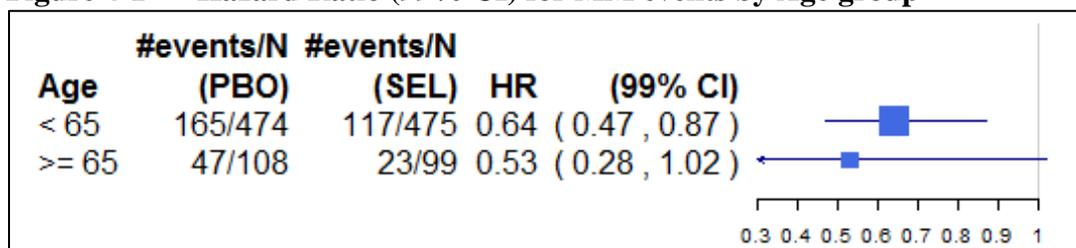
### 4.1 Age, Gender, and Race group

The following subsections present the incidence of the primary composite endpoint and the estimate of treatment effect in the subgroups of the gender, age and race.

#### 4.1.1 AGE

The majority of subjects are under 65 years old (82.1%). There are only about 1% subjects are over 75 years old [Table 3-3], so the following analysis grouped 65-74 years old with over 75 years old together into one group. The observed treatment effects were consistent with the primary finding across both age groups [Figure 4-1].

**Figure 4-1 Hazard Ratio (99% CI) for MM events by Age group**

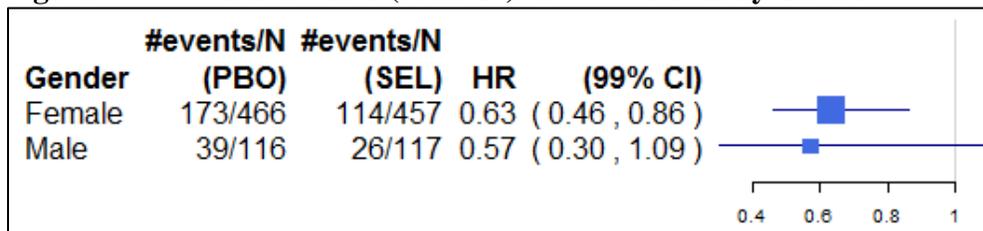


[Source: Reviewer’s Results]

### 4.1.2 GENDER

The consistent findings as the primary efficacy results are observed within both gender group. Within male and female subjects, selexipag group experienced fewer proportion of CEC-confirmed MM events than placebo group, see Figure 4-2.

**Figure 4-2 Hazard Ratio (99% CI) for MM events by Gender**

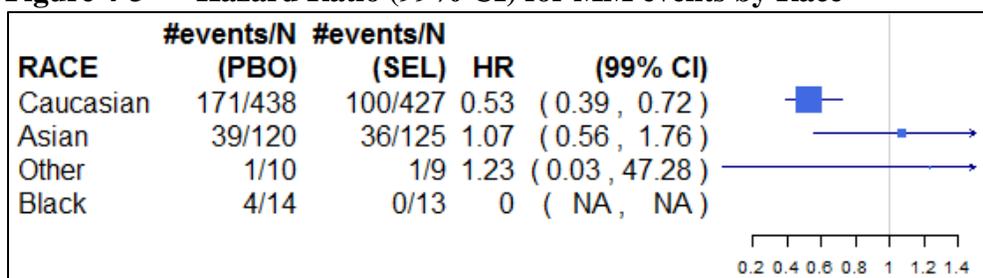


[Source: Reviewer’s Results]

### 4.1.3 RACE

The majority of patients was Caucasian (65%) and with about 21% of Asian subjects. There were very few Black/African American subjects. For the Asian race, selexipag group almost observed an identical proportion of CEC-confirmed MM events as placebo group. The hazard ratio between two treatment groups was 1.07, see Figure 4-3.

**Figure 4-3 Hazard Ratio (99% CI) for MM events by Race**

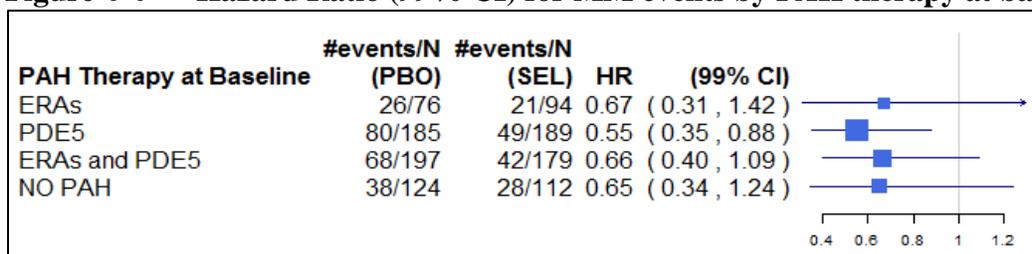


[Source: Reviewer’s Results]

## 4.2 Other Subgroup Populations

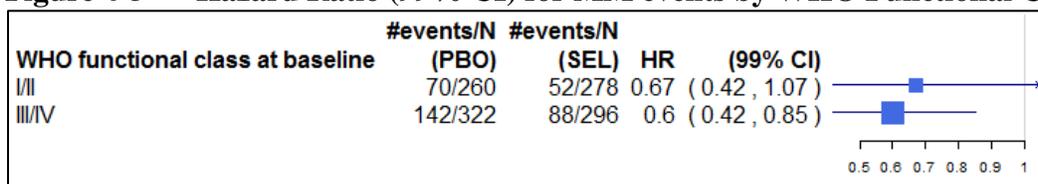
The occurrence of a first MM event confirmed by the CEC was further summarized for the following additional subgroups: PAH therapy at baseline [Figure 4-4], NYHA/WHO FC at baseline [Figure 4-5] and geographical regions [Figure 4-6]. There were no notable deviants from the primary findings within each subgroup, except the Asia region. In the subgroup of Asia (geographical region), the hazard ratio for selexipag versus placebo was 1.02.

**Figure 4-4 Hazard Ratio (99% CI) for MM events by PAH therapy at baseline**



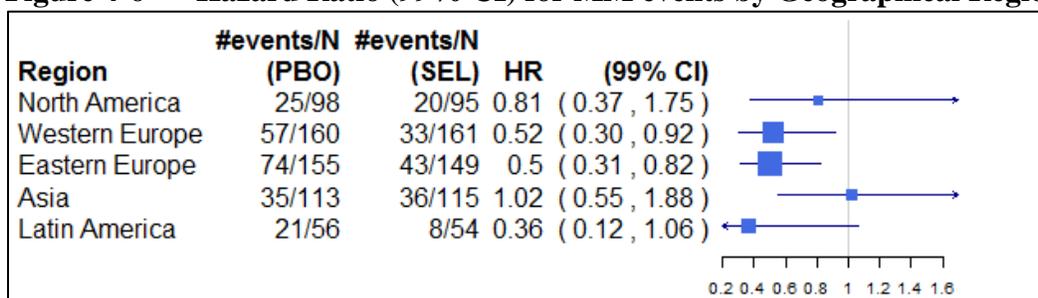
[Source: Reviewer’s Results]

**Figure 4-5 Hazard Ratio (99% CI) for MM events by WHO Functional Class at baseline**



[Source: Reviewer’s Results]

**Figure 4-6 Hazard Ratio (99% CI) for MM events by Geographical Regions**



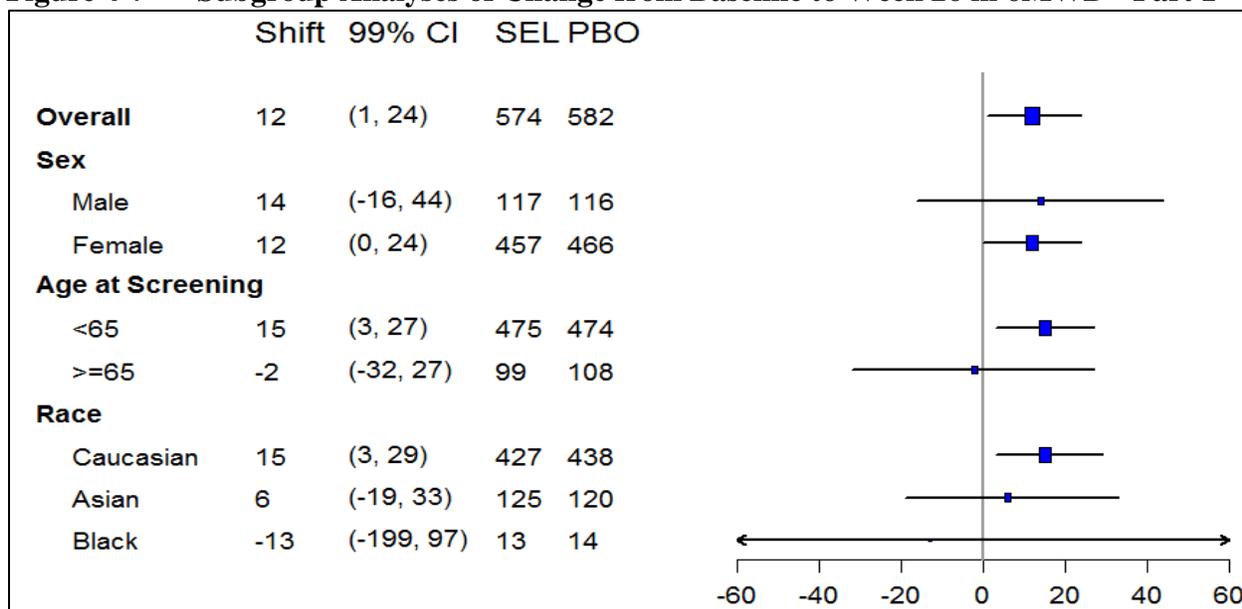
[Source: Reviewer’s Results]

### 4.3 Subgroup Analyses on change from Baseline to Week 26 in 6MWD

The secondary endpoint of absolute change from Baseline to Week 26 in 6MWD at trough showed a statistically significant difference favoring selexipag over placebo [Section 3.2.5.2]. The subgroup analyses of this secondary endpoint are provided in Figure 4-7 and Figure 4-8. The observed treatment effect, location shift in Hodges-Lehmann estimator, on the 6MWD at Week 26 was consistent across the most subgroups. The median absolute changes from baseline to Week 26 in 6MWD measured at trough in the selexipag group were either neutral or inferior when compared to the placebo group in the following subgroups:

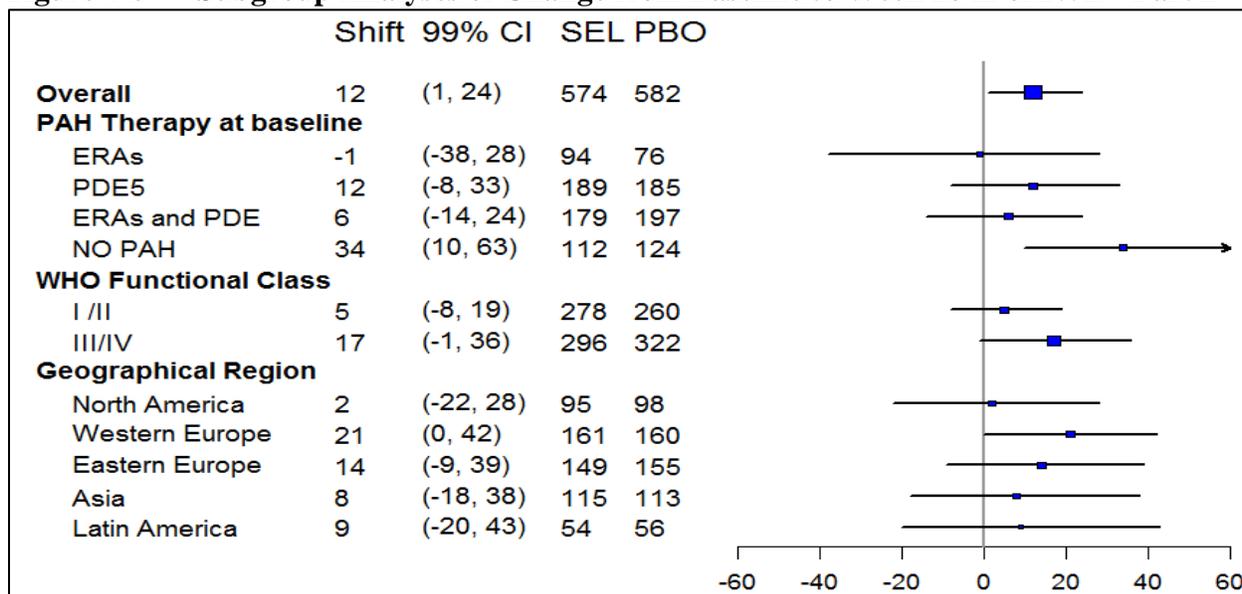
- patients who are over 65 years of age
- patients who had ERAs therapy at baseline
- Black
- North America

**Figure 4-7 Subgroup Analyses of Change from Baseline to Week 26 in 6MWD - Part 1**



[Source: Reviewer’s Results]

**Figure 4-8 Subgroup Analyses of Change from Baseline to Week 26 in 6MWD - Part 2**



[Source: Reviewer’s Results]

## 5 SUMMARY AND CONCLUSIONS

### 5.1 Statistical Issues and Collective Evidence

There were a number of major statistical changes throughout six major protocol amendments. The following are the summary of these changes:

- The primary endpoint was modified a couple of times and finally settled as MM event in the Amendment 6.

- The number of primary events was increased in Amendment 4 due to changes in the sample size calculation assumption. This amendment 4 also proposed addition of an interim analysis.
  - In order to eliminate any concern that Amendment 4 could be considered informative, the primary events observed until 16 August 2011 are censored in the primary analysis.
- All above changes have occurred before any version of Statistical Analysis Plans.

Study GRIPHON employed a group-sequential design for the primary efficacy endpoint with options to recommend stop for futility or for compelling and robust efficacy at the interim analysis. In case of rejection of the null hypothesis in the primary statistical analysis of the primary efficacy endpoint, the null hypotheses for the secondary efficacy endpoints were tested in a conditional hierarchical manner. A null hypothesis was rejected if and only if the main analysis of the endpoint and all main analyses of preceding secondary efficacy endpoints resulted in rejection of respective null hypotheses.

In the interim analysis, selexipag did not provide compelling evidence to stop the trial for overwhelming efficacy. However, we can argue it almost made the stopping criterion. In the Final Analyses, the main analysis of CEC-confirmed MM event up to 7 days after last study drug intake in the treatment period was recorded for 140 patients in the selexipag group compared to 212 patients in the placebo group. In the final main time-to-event analysis the 1-sided unstratified log-rank p-value was  $< 0.0001$ . The hazard ratio for selexipag versus placebo for the occurrence of an MM event was 0.61 (99% CI: 0.46, 0.81). The treatment difference in the primary endpoint in GRIPHON was driven by hospitalization due to PAH and the composite component of disease progression, while there was a higher proportion of patients with death as the first MM event in the selexipag group (4.4% vs 2.7% in the placebo group).

The secondary endpoint of change from baseline to Week 26 in 6MWD measured at trough showed a statistically significant difference favoring selexipag over placebo. The treatment effect (location shift using Hodges-Lehmann method) versus placebo in the selexipag group was 12 m (99% CI: 1, 24) with 1-sided Wilcoxon-Mann-Whitney  $p = 0.0027$ . However, both treatment groups deteriorated in term of 6MWD. Therefore, the significant finding in 6MWD has little or no clinical meaning.

The secondary endpoint of absence of worsening in NYHA/WHO FC from baseline to Week 26 did not show a difference between selexipag and placebo. The common odds ratio for selexipag versus placebo was 1.161 with 99% CI (0.811, 1.664). Hence, according to the statistical analysis plan, this review ended the statistical analyses at the endpoint of absence of worsening in NYHA/WHO FC.

## 5.2 Conclusions and Recommendations

The efficacy of selexipag in the long-term treatment of Pulmonary Arterial Hypertension (PAH) is based on the data from the pivotal Phase 3 study AC-065A302 (GRIPHON). GRIPHON is the largest clinical outcome study in PAH conducted to date. It was adequately designed as a Morbidity/Mortality (MM) event-driven clinical outcome trial with strict statistical specifications.

The primary objective of the study was met. A statistically highly significant 39% risk-reduction for the occurrence of a first MM event up to end of treatment (EOT) + 7 days was demonstrated

with selexipag treatment. The MM event was defined by the first event of death (all causes), hospitalization for PAH worsening, lung transplantation, atrial septostomy, initiation of parenteral prostanoids or chronic oxygen therapy, or disease progression, all adjudicated by an independent CEC blinded to treatment allocation. This represents a clinically highly relevant treatment-effect in a progressive and ultimately fatal cardiovascular disease.

This review does not consider the statistically significant difference between selexipag and placebo in the 6-minute walk distance, secondary symptomatic variable, has any clinical relevance.

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

/s/  
-----

STEVE G BAI  
07/29/2015

HSIEN MING J HUNG  
07/29/2015

## STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

**NDA Number: 207947**

**Applicant: Actelion**

**Stamp Date: 12/17/2014**

**Drug Name: UPTRAVI(Selexipag) NDA/BLA Type: 505(b)(1)**

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

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

**IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE?**   Yes  

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

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

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

Steve Bai

2/2/2015

\_\_\_\_\_  
Reviewing Statistician

\_\_\_\_\_  
Date

Hsien Ming Hung

2/2/2015

\_\_\_\_\_  
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/  
-----

STEVE G BAI  
02/02/2015

HSIEN MING J HUNG  
02/02/2015