CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:
22-512

STATISTICAL REVIEW(S)

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 #/Serial #:	22-512
DRUG NAME:	Dabigatran Etexilate Mesylate
INDICATION:	Stroke Prevention in Atrial Fibrillation
APPLICANT:	Boehringer Ingelheim Pharmaceuticals Inc.
DATE OF RECEIPT:	12/15/2009
REVIEW PRIORITY:	Standard
BIOMETRICS DIVISION:	Division of Biometrics I
STATISTICAL REVIEWER:	Steve Bai, Ph.D.
CONCURRENT REVIEWER:	James Hung, Ph.D. Director DBI
MEDICAL DIVISION:	Division of Cardiovascular and Renal Products
CLINICAL TEAM:	Aliza Thompson, MD
	Bach Nhi Beasley, MD
PROJECT MANAGER:	Alison Blaus

Table of Contents

List of Tables

List of Figures

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Overall, RE-LY demonstrated that both doses of dabigatran were non-inferior to warfarin and DE 150 was superior to warfarin for the primary (stroke/SEE) efficacy endpoints. Furthermore, the secondary (stroke/SEE/death and stroke/SEE/PE/MI/vascular death) efficacy endpoints also met the above claims numerically. However, sponsor did not specify the statistical testing rules and margins for these endpoints in the TSAP. Therefore, these findings can only be viewed as exploratory findings.

There was no discrepancy results found in any of the sensitivity analyses. Although, DE 150 did not show superiority for US subjects statistically, but it was still non-inferior to warfarin and the point estimate (hazard ratio) also less than 1.00. All the subgroup analyses performed in Section 4 were consistent with the primary efficacy results. Hence, RE-LY's finding is very robust. Furthermore, based on the reviewer's analysis on the impact of different end of trial dates, the dabigatran doses achieved the non-inferiority long before the end of trial date and DE 150 achieved superiority to warfarin more than one year before the end of trial date, see Figure 3.3.

1.2 Brief Overview of Clinical Study

RE-LY is a randomized, parallel group, active-controlled, non-inferiority trial of 2 blinded doses of Dabigatran Etexilate compared with open-label warfarin in patients with non-valvular AF. The trial was designed to evaluate whether 110 mg bid and 150 mg bid of Dabigatran Etexilate are non-inferior to adjusted dose warfarin in the prevention of stroke and systemic embolism in non-valvular AF patients with at least 1 additional risk factor for stroke. A total of 18,113 subjects (1:1:1) were randomized and the total number of subjects with adjudicated stroke/SEE was 513.

1.3 Statistical Issues and Findings

<u>The non-inferiority margin</u>

The proposed testing hypothesis for RE-LY was that whether either dabigatran doses (110 mg and 150 mg) were non-inferior to warfarin in reducing the incidences of Stroke/SEE. The non-inferiority margin of 1.46 for the hazard ratio in the sponsor's study report was derived based on the historical placebo controlled trials using the 95%-95% rule. This rule utilized the lower bound of the 95% confidence interval of the hazard ratio for warfarin versus placebo for the derivation of the non-inferiority margin, and the upper bound of the 95% confidence limit for dabigatran versus warfarin for the statistical test. The margin 1.46 used in the study design preserved at least 50% of warfarin's effect on the risk ratio scale using the lower bound of the 95% of the risk ratio of placebo over warfarin. However, a smaller margin of 1.38, derived to preserve the effect of warfarin on the Log scale, was recommended by a regulatory agency. In spite of this discrepancy on the margin, both dabigatran doses were non-inferior to warfarin based on the sponsor's efficacy findings.

<u>Summary of the historical trials and constancy assumption</u>

The effectiveness of warfarin has been studied both in placebo-controlled and active-controlled trials. There are six placebo-controlled studies of warfarin involved the patients with AF between 1989 and 1992. All these trials showed a consistent efficacy for warfarin in preventing stroke and other cardiovascular events, despite differences in their designs and patient populations. The primary outcomes of these studies are summarized in Table 10. Almost all of the trials showed significant reduction in the primary endpoint event by warfarin against placebo. Trial CAFA failed to show a significant benefit over placebo, but the estimated warfarin effect from this trial was consistent with those observed from the other trials.

Even if the historical studies are consistent, a critical consideration in deciding upon the NI margin derived from these studies is whether the constancy assumption is reasonable. To evaluate the plausibility of this constancy assumption, one might compare some features of the six placebo-controlled warfarin studies with the RE-LY study. There is considerable heterogeneity in the demographic characteristics of these studies. The draft guidance listed number of characteristics, such as a history of stroke or TIA, see Table 11. The most of characteristics are similar among the historical studies with RE-LY, but the history of stroke or TIA and CAD are much higher in RE-LY; see Table 11 on page 23 of this review.

Increase of sample size

The study was originally designed as an event driven trial. Based on an estimated yearly event rate of 1.6% and a two-year enrolment period and one-year follow up, a total of 15,000 subjects were planned to be randomized from approximately 800 centers. Due to rapid enrollment, 15,000 subjects were randomized in 1.5 years (18 months). Sponsor claimed that if the recruitment was stopped at that time, the last randomized subjects would have had to follow up for more than 1 year to achieve the planned total number of events, if the actual event rate was as expected. In addition, sponsor also claims that the actual event rate could be less than 1.6% based on other published studies. Therefore, sponsor decided to continue the recruitment as planned, which resulted a total of 18,113 subjects were randomized and the total number of subjects with adjudicated stroke/SEE was increased to 519. The above changes were added to protocol's second amendment on May of 2007. In order to validate the final primary efficacy results, both the sponsor and this reviewer had performed the sensitivity analysis for the first 450 adjudicated primary events. Based on all the analyses results, both doses of dabigatran have met the pre-specified non-inferior margin to warfarin to conclude that the two doses are effective for the stroke prevention in AF patients. Furthermore, DE 150 was superior to warfarin as well, see Table 4.

2 INTRODUCTION

2.1 Overview

Atrial Fibrillation is the most common sustained cardiac rhythm disturbance. The prevalence of paroxysmal or persistent AF is estimated at 0.4% of the general population, including up to 1% of all adults. The prevalence of AF increases with age. It occurs in less than 1% of those under 60 years of age but in more than 6% of those over 80 years of age. AF has significant morbidity, mortality, and economic cost, due to the occurrence of both hemodynamic impairment and thromboembolic events. The hemodynamic impairment and rhythm disturbances may be

symptomatic and can lead to a decrease in quality of life. However, most of the mortality and functional impairment associated with AF is due to the occurrence of ischemic stroke and systemic emboli. AF patients also have concomitant coronary artery disease, for which they should normally receive acetylsalicylic acid (ASA). However due to a higher rate of bleeding when anticoagulants and ASA are co-administered, one of these agents may either be withheld or dose-adjusted in such patients.

The vitamin K antagonists (VKAs, coumadins), typified by warfarin, are the most widely prescribed oral anticoagulants. In several adequate and well-controlled trials, warfarin decreased the risk of stroke/systemic thromboembolism by 68% versus placebo. This class of drugs when used in patients with AF also has shown to have a higher risk of bleeding at therapeutic doses than ASA alone. VKAs have a slow onset and offset of action, high inter- and intra-individual variability in their effective plasma concentrations, and have a high potential for food and drug interactions.

Dabigatran Etexilate is the orally bioavailable prodrug of Dabigatran, a novel thrombin inhibitor. Dabigatran Etexilate, a prodrug, does not have any antithrombin activity.

2.2 Data Sources

The sponsor's submitted data are stored in the following directory of the CDER's electronic document room: \\Cdsesub1\evsprod\NDA022512

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

The following description is based on the sponsor's clinical study report. Any discrepancy between the study report and study protocol will be discussed in the section of statistical reviewer's comments.

3.1.1 RE-LY STUDY

RE-LY is a randomized, parallel group, active-controlled, non-inferiority trial of 2 blinded doses of Dabigatran Etexilate compared with open-label warfarin in patients with non-valvular AF. The trial was designed to evaluate whether 110 mg bid and 150 mg bid of Dabigatran Etexilate are non-inferior to adjusted dose warfarin (target INR of 2.0 to 3.0) in the prevention of stroke and systemic embolism in non-valvular AF patients with at least 1 additional risk factor for stroke.

Objectives

The primary objective is to demonstrate that the efficacy and safety of 2 blinded doses (110 mg bid and 150 mg bid) of Dabigatran Etexilate are non-inferior to adjusted dose warfarin (target INR 2-3) for the prevention of stroke and systemic embolism in subjects with non-valvular AF with at least 1 additional risk factor for stroke.

Study Design

This was a Prospective Randomized Open trial with Blinded outcome Evaluation (PROBE) study with 2 doses of Dabigatran Etexilate (110 mg bid, 150 mg bid) compared to adjusted warfarin therapy, INR 2.0-3.0. Approximately 6,000 subjects per treatment group were randomized over 2 years with a further year of follow-up to a common termination.

The trial was conducted from December 22, 2005 to March 15, 2009. There were 1,044 sites selected from 44 countries and 951 sites randomized at least 1 subject. The duration of treatment was expected to be a median of 20-24 months, with a minimum of 12 months' treatment after the last subject was randomized and a maximum treatment of approximately 3 years.

There were 5 protocol amendments written for this study. Amendment 1 mandated balanced randomization of warfarin-naïve and warfarin –experienced subjects at each site. In order to obtain balanced cohorts of both VKA-experienced and –naïve subjects, investigational sites were expected to recruit both types of subjects. With rapid recruitment of predominantly VKA-experienced subjects (80%) in the first 7 months of the trial, Amendment 1 (dated 31 Aug 2006) was implemented to ensure that balanced cohorts were recruited. The definition of VKA-naïve was expanded from 1 month to 2 months or less of lifetime VKA use. Amendment 2, dated 24 May 2007, increased the target sample size to 18,000 from originally proposed 15,000. The 15,000 patients were planned based on a two-year enrollment and one year of follow-up and a yearly event rate of 1.6%. Due to the faster enrollment, 15,000 patients will be randomized prior to the planned date. In order to maintain the statistical power in case of event rate < 1.6% within the original study time line, the enrollment should continue as planned. It is predicted that the number of patients randomized will be increased from 15,000 to 18,000. Amendment 3, 4 and 5 did not have any statistical issues.

The logistic of a double-blind study design employing warfarin, which is frequently monitored and dose-adjusted, compared with Dabigatran, which is neither monitored nor dose-adjusted, and are complex. A dummy INR monitoring system, with an algorithm for generating false INRs for Dabigatran subjects would need to be established, further complicating recruitment of both centers and subjects. This trial used the Prospective Randomized Open trial with Blinded Evaluation of outcomes (PROBE) design. A key element of the PROBE design was to use blinded adjudicators to reduce potential bias in the evaluation and classification of important study outcome events. The following measures were used to decrease open-label biases:

- Blinded Adjudication of events by at least 2 independent adjudicators
- Database and data handling assigned to an academic group independent from the sponsor
- Blinding of sponsor and trial management personnel to "by treatment" analyses during trial
- Oversight by DSMB
- CRF construction to elicit events based on investigations and other assessments performed by the site.

Efficacy Measures

The primary endpoint for this study is the incidence of stroke (including hemorrhagic) or non-Central Nervous System (CNS) systemic embolism, hereafter referred to as systemic embolism.

The secondary endpoints are:

- incidence of stroke (including hemorrhagic), systemic embolism, all death
- incidence of stroke (including hemorrhagic), systemic embolism, pulmonary embolism, acute myocardial infarction, or vascular deaths (includes deaths from bleeding)

There are two other efficacy endpoints:

- individual or composite occurrences of ischemic stroke (fatal and non-fatal), systemic embolism, pulmonary embolism, acute myocardial infarction, TIAs, vascular death (includes deaths from bleeding), all deaths, and hospitalizations
- Net Clinical Benefit (NCB) as measured by the composite of the clinical endpoint of stroke, systemic embolism, pulmonary embolism, acute myocardial infarction, all cause deaths, and major bleeds.

Statistical Hypotheses

The null hypothesis was that hazard ratio of Dabigatran vs. warfarin was larger than or equal to the specified non-inferiority margin $\delta = 1.46$. The alternative hypothesis was that the hazard ratio was less than 1.46. Since there were two comparisons of Dabigatran vs. warfarin, the Hochberg procedure was used to handle the multiple comparisons. To use the Hochberg procedure, the Dabigatran dose with the largest hazard ratio vs. warfarin were to be tested first for non-inferiority at $\alpha=0.025$ (one-sided) level. If the non-inferiority would be concluded from this comparison, then the non-inferiority vs. warfarin for both Dabigatran doses would be claimed. Otherwise, the non-inferiority for this dose would not be claimed and the other Dabigatran dose were to be compared to warfarin at $\alpha=0.0125$ (one-sided) level for non-inferiority.

As specified in the Trial Statistical Analysis Plan (TSAP), superiority testing was to be performed to compare Dabigatran to warfarin for the primary endpoint when the non-inferiority claim was established.

Efficacy Analysis

The primary analysis was performed by using the randomized set, which included all randomized subjects in the treatment groups to which they were randomized, regardless of whether the subjects took randomized study medication or not. The time to the occurrence of the primary endpoint event was computed as (event date – randomization date) +1. Subjects who did not have primary endpoint events during the trial period were considered to be censored. The time to censoring was computed as (study termination date – randomization date) + 1.
The yearly event rate for treatment group was computed as the total number of events that occurred in that treatment group divided by the total subject exposure in years (subject years) in that group. For a given subject, exposure was computed from the date of randomization to the date of study termination, using the randomized set.

The primary analyses include the following: yearly event rate summaries, Kaplan-Meier curves and Cox regression analyses. All secondary outcomes were analyzed using the Cox regression model with treatment as the factor in the model.

Sensitivity Analyses

An analysis of the primary endpoint including only the first 450 adjudicated primary endpoint events was performed as a sensitivity analysis since the originally planned number of events was 450. Subjects without primary events were censored at the onset date of the 450th primary event, or study termination date which ever occurred earlier.

Another analysis for the primary endpoint was performed by including all subjects randomized to dabigatran treatment and subjects randomized to warfarin who achieved good INR control, such as >= 65% of time INR in range 2-3 during the treatment period.

Overall, 8,542 (47%) subjects completed the trial without any interruption, 2,736 (15%) subjects permanently discontinued their study medication. Lastly, 6,762 subjects had a temporary interruption of study medication. Therefore, number of different on-treatment analyses by recoding event status and time to outcomes for those temporary discontinued subjects are included in this review. This review included three different recoding schemes: (1) censoring at first discontinuation of study medication, (2) censoring at last study medication date, and (3) censoring at 7 days after first discontinuation of study medication. The statistical analyses will be same as the primary efficacy analysis.

Sample Size Considerations

The RE-LY assumed a yearly event rate of 1.6% for both Dabigatran and warfarin, with 5,000 subjects per treatment group to be recruited in 2 years and followed up for 1 additional year to achieve 150 events per treatment group. Within these parameters, each comparison had approximately 90% power to conclude the non-inferiority of Dabigatran to warfarin at a one-sided α=0.025 level based on the derived non-inferiority margin of 1.46. With a total of 15,000 subjects randomized to the 2 Dabigatran doses and warfarin at a 1:1:1 ratio, to achieve a total of 450 events, using the Hochberg procedure to compare each Dabigatran dose to warfarin, the trial had approximately 84% power to conclude the non-inferiority of both Dabigatran doses to warfarin using the non-inferiority margin of 1.46.

A total of 15,000 subjects were recruited in less than 2 years (18 months). If the recruitment was stopped at that time, the last randomized subject would have had to be followed up for more than 1 year to achieve the planned total number of events, if the actual event rate was as expected. In addition, based on the results from other published studies, the actual event rate could be less than 1.6%. Because of these concerns, the operational committee decided to continue the recruitment as planned. As a result, a total of 18,113 subjects were randomized. It was expected that if the actual event rate was as planned, the statistical power would be increased.

3.1.1.1 Patient Disposition, Demographic and Baseline Characteristics

In general, there were no large differences among the three treatment groups in subject baseline demographic and disease characteristic information. Detailed baseline demographic and disease characteristics are presented in Table 1.

Table 1 Baseline Demographic Information

	DE 110 mg	DE 150mg	Warfarin	Total
Randomized [N]	6015	6076	6022	18,113
Age (mean, years)	71.4	71.5	71.6	71.5
Male (%)	64.3	63.2	63.3	63.6
Race: white (%)	70	70.2	69.8	70
Weight (mean, Kg)	82.9	82.4	82.6	82.6
VKA naïve (%)	50	49.8	51.4	50.4
Never on VKA (%)	31.1	31.4	32.7	31.7
CrCL (median, ml/min)	68.7	67.9	68.5	68.4
Systolic BP (mean, mmHg)	130.8	130.9	131.2	131
Diastolic BP (mean, mmHg)	77	77	77.1	77
AF type [N(%)]				
Persistent	1950 (32.4)	1909 (31.4)	1930 (32.0)	5789 (32.0)
Paroxysmal	1929 (32.1)	1978 (32.6)	2036 (33.8)	5943 (32.8)
Permanent	2132 (35.4)	2188 (36.0)	2055 (34.1)	6375 (35.2)
Previous cardioversion	1658 (27.6)	1683 (27.7)	1651 (27.4)	4992 (27.6)
Previous AV nodal ablation	119 (2.0)	136 (2.2)	132 (2.2)	387 (2.1)
Pacemaker	613 (10.2)	679 (11.2)	646 (10.7)	1938 (10.7)
Implantable defibrillator	136 (2.3)	138 (2.3)	125 (2.1)	399 (2.2)
Regions [N(%)]				
USA, Canada	2166(36.0)	2200(36.2)	2167(36.0)	6533(36.1)
Central Europe	707(11.8)	706(11.6)	706(11.7)	2119(11.7)
Western Europe	1544(25.7)	1555(25.6)	1552(25.8)	4651(25.7)
Latin America	320(5.3)	320(5.3)	316(5.2)	956(5.3)
Asia	923(15.3)	933(15.4)	926(15.4)	2782(15.4)
Other	355(5.9)	362(6.0)	355(5.9)	1072(5.9)

3.1.1.2 Primary Efficacy Results

First of all, the results presented in this review were all this reviewer's own results. Furthermore, they also confirmed the sponsor's results. The primary objective in this study was to determine if Dabigatran was non-inferior to warfarin in reducing the occurrence of the composite endpoint, stroke/SEE. Comparisons between treatment groups for stroke/SEE were performed using a Cox regression analysis with treatment in the model. Descriptive statistics, such as event numbers and Kaplan-Meier plots, are also presented in Table 2.

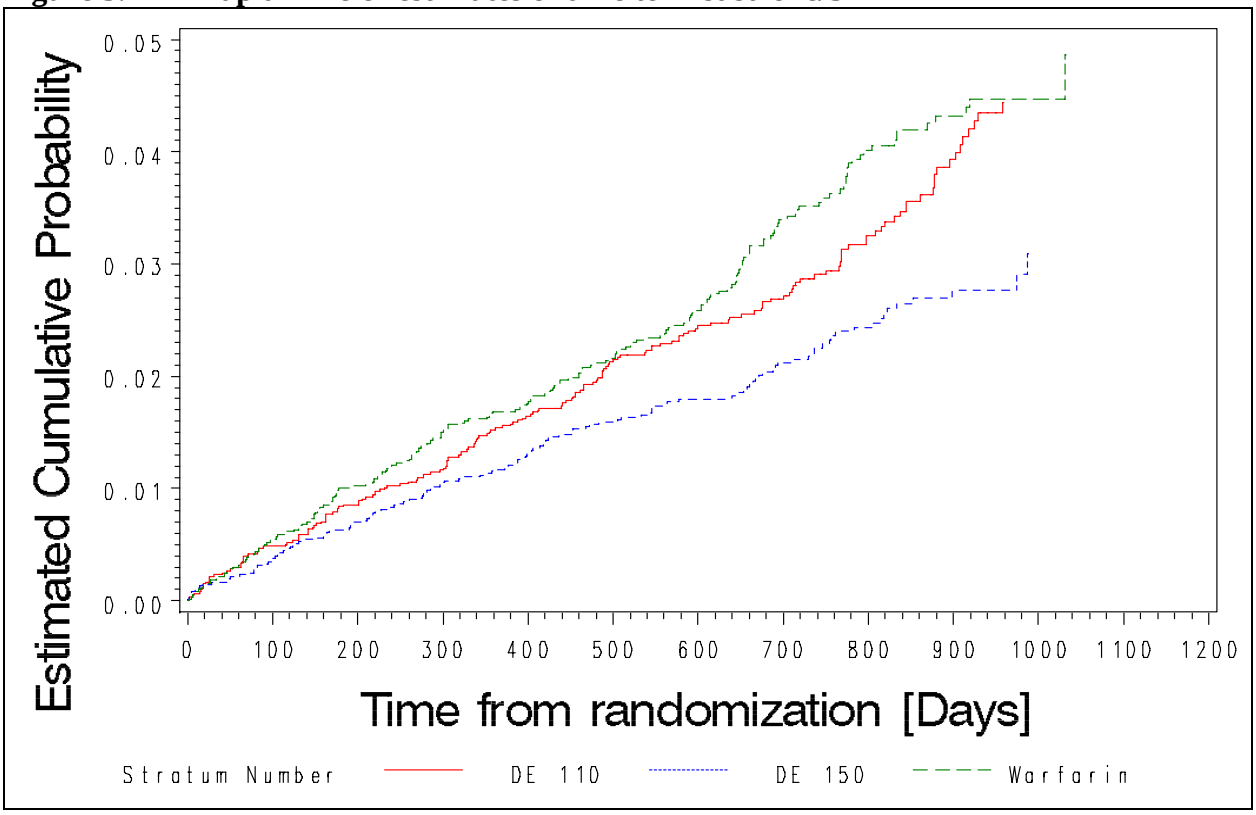
Table 2 Frequency for stroke/SEE in randomized set

	DE 110 mg	DE 150 mg	Warfarin
Subjects randomized	6015	6076	6022
Subjects with stroke/SEE	183	134	202
Stroke	183	132	194
Ischemic stroke	162	111	139
Haemorrhagic stroke	14	12	45
Stroke of uncertain	7	9	10
SEE	15	14	21

[Source: reviewer's results]

A total of 519 adjudicated first stroke/SEEs were observed during the trial: 183, 134 and 202 events in the DE 110, DE 150 and warfarin groups, respectively (Table 2) The Kaplan-Meier estimates are shown in Figure 3.1.

Figure 3.1 Kaplan-Meier estimates of time to first stroke/SEE



[Source: Reviewer's Results]

Non-inferiority of both dabigatran doses compared to warfarin was demonstrated. The hazard ratio for stroke/SEE of DE 110 over warfarin was 0.90, with the 95% confidence limits (CI) of (0.74, 1.10). The upper bound of the 95% CI is below 1.46, the protocol specified margin, for both doses. Relative risk reductions for stroke/SEE by DE 110 and DE 150 were 10% and 35%, respectively, in comparison to warfarin. Furthermore, DE 150 was superior to warfarin for the

primary endpoint of stroke/SEE. The hazard ratio of DE 150 over warfarin was 0.65, with the 95% CI of (0.52, 0.81). The upper bound of the 95% CI is below 1.00, See Table 3.

Table 3 Hazard ratios and CIs for stroke/SEE, randomized set.

	DE 110 mg vs. Warfarin	DE 150 mg vs. Warfarin
#Events/N	183/6015 vs. 202 /6022	134/6076 vs. 202/6022
Hazard ratio (SE)	0.90 (0.09)	0.65 (0.07)
95% CI	0.74, 1.10	0.52, 0.81
P-value for NI using 1.46	0.0001	0.0001
P-value for superiority	0.2943	0.0001

[Source: Reviewer's results]

3.1.1.3 Sensitivity analyses for stroke/SEE

Sensitivity analyses were performed for the primary endpoint in order to provide evidence that the primary analysis is robust. The following sensitivity analyses are presented in this section:

1. analyses of the first 450 adjudicated events;
2. analyses of all dabigatran and warfarin subjects with INR in 2-3 ≥65% of the time and <65% of the time;
3. analyses of as-treated subjects.

Analysis of the first 450 adjudicated events

The original targeted number of events for stroke/SEE in this study was 450. At the end of the study, 519 adjudicated stroke/SEEs were reported. The analysis including the first 450 events was performed as a sensitivity analysis. The 450th adjudicated stroke/SEE occurred on October 30, 2008. Subjects without a stroke/SEE were considered censored on this date for purposes of this analysis.

Table 4 Hazard ratios and CIs for stroke/SEE randomized set, data cutoff at the event onset date of the 450th adjudicated event

	DE 110 mg vs. Warfarin	DE 150 mg vs. Warfarin
#Events/N	159/6015 vs. 170/6022	121/6076 vs. 170/6022
Hazard ratio	0.936	0.70
95% CI	0.75, 1.16	0.56, 0.89

[Source: Reviewer's results]

As in the primary analysis, both doses of dabigatran were non-inferior to warfarin, and DE 150 mg was superior to warfarin, see Table 4.

Analyses by INR control

The subjects on warfarin had their INR level measured throughout the whole trial and the mean percent of time of INR in 2-3 were computed for each warfarin subject as well. Hence, the results of the sensitivity analyses of all dabigatran and warfarin subjects with INR in 2-3 ≥ 65% of the time and < 65% of the time for the primary endpoint in provided in Table 5.

Table 5 Hazard ratios and 95% CI for stroke/SEE by INR control for warfarin

Mean % of the time of INR in range 2-3 ≥ 65%		
	DE 110 mg vs. Warfarin	DE 150 mg vs. Warfarin
#Events/N	183/6015 vs. 89/3195	134/6076 vs. 89/3195
Hazard ratio	1.12	0.81
95% CI	0.87, 1.44	0.62, 1.05
Mean % of the time of INR in range 2-3 < 65%		
	DE 110 mg vs. Warfarin	DE 150 mg vs. Warfarin
#Events/N	183/6015 vs. 113/2827	134/6076 vs. 113/2827
Hazard ratio	0.73	0.53
95% CI	0.58, 0.92	0.41, 0.67

[Source: Reviewer's results]

Results of the above sensitive analysis showed that the non-inferiority of both dabigatran doses compared to warfarin for stroke/SEE is maintained compared to well-controlled warfarin subjects when using a NI margin of 1.46. Superiority of both dabigatran doses compared to warfarin is demonstrated when dabigatran subjects are compared to subjects on warfarin whose mean percent of time of INR in 2-3 was <65% since the upper bound of both hazard ratio comparisons are below 1.00.

On-Treatment Analysis

During the further examination of the sponsor's datasets, the reviewer has noticed that around 13,151 subjects had a temporary interruption of study medication among all three treatment groups during the course of the trial. Some of them went back in a few short periods or longer periods. And others never went back to their assigned treatment.

Table 6 Hazard ratios and CIs for stroke/SEE, as-treated set.

Censoring Scheme	**DE 110 vs warfarin**		**DE 150 vs warfarin**	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*
Censoring at first discontinuation of study medication (temporary or permanent)	0.86 (0.59, 1.27)	0.45	0.56 (0.40, 0.86)	0.009
Censoring at last study medication date	0.70 (0.53, 0.92)	0.01	0.45 (0.32, 0.61)	<.0001
Censoring at 7 days after first discontinuation of study medication (temporary or permanent)	0.85 (0.64, 1.13)	0.273	0.62 (0.46,0.85)	0.0028

[Source: Reviewer's results. *p-value is for superiority]

The reviewer, hence, conducted the following as-treatment analyses by re-code the time to censoring: (1) censoring at first discontinuation of study medication, (2) censoring at last study medication date, and (3) censoring at 7 days after first discontinuation of study medication. The detailed recoding mechanism for analysis (1) is described as the following: the data is re-coded based on their first discontinuation date (FDdate). For the censored subjects, if their FDdate occurred prior to their study termination date, then the time to censoring will be recoded as FDdate – Randomization date +1. For the event subjects, if their FDdate occurred prior to their event date, then the time to event will be recoded as FDdate – Randomization date +1 and the events will be changed to the censors. The recoding mechanism for analyses (2) and (3) would

be same as (1). Based on the findings in Table 6, the results are consistent with the primary efficacy analysis.

3.1.1.4 Secondary Efficacy Analysis

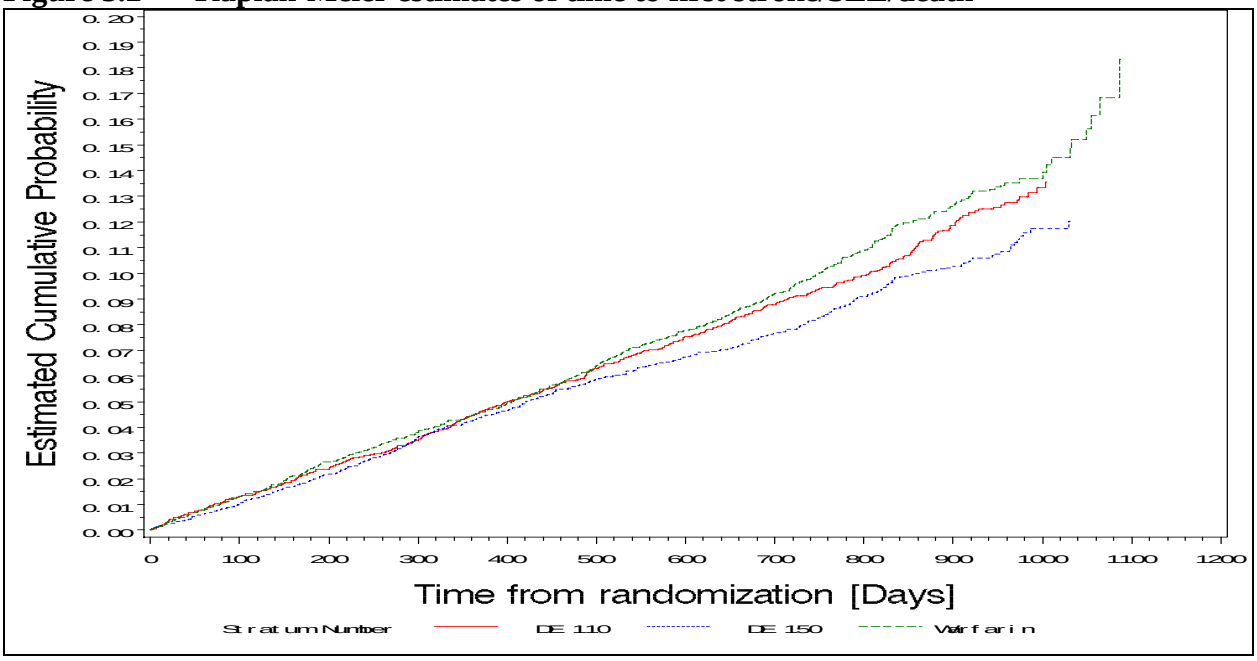
There were two secondary endpoints were specified in the protocol: 1) composite of stroke, SEE and all cause death, and 2) composite of stroke, SEE, PE, MI and vascular death.

Analysis of stroke, SEE, and all cause death

A total 1,710 stroke/SEEs/all cause deaths were observed during the trial: 577, 520 and 613 from the DE 110, DE 150 and the warfarin groups, respectively. The yearly event rate for the composite endpoint was the lowest in the DE 150 group (4.85%, 4.32% and 5.20%) in the DE 110, DE 150 and warfarin groups, respectively).

The Kaplan-Meier estimate for stroke/SEE/death shows the separation among the three curves as 500 days after the date of the randomization, with the DE150 group starting to be lower than the warfarin curve. DE 150 then has the lowest occurrence rate of death, with DE 110 also consistently below warfarin after about 1 year and through the end of the study, see Figure 3.2.

Figure 3.2 Kaplan-Meier estimates of time to first stroke/SEE/death



[Source: reviewer's result]

The risk reduction for the DE 110 group in stroke/SEE/death was 7% in comparison to warfarin, which was not statistically significant. The relative risk reduction for the DE 150 group was 17%, which was significant (p-value =0.0015) (Table 7).

Table 7 Hazard ratios and 95% CI for composite endpoint of stroke/SEE/death

	DE 110 mg vs. Warfarin	DE 150 mg vs. Warfarin
#Events/N	575/6015 vs. 609/6022	518/6076 vs. 609/6022
Hazard ratio (SE)	0.93 (0.05)	0.83 (0.05)
95% CI	0.83, 1.05	0.74, 0.93
P-value	0.2206	0.0015

[Source: reviewer's results]

Analysis of stroke, SEE, PE, MI and vascular death

For the other secondary composite endpoint (stroke/SEE/PE/MI/vascular death), results followed the same pattern as composite endpoints stroke/SEE and stroke/SEE/death. The event rates for DE 110 and warfarin were similar, while the event rate in the DE 150 was lower. A total 1,435 such composite endpoints were observed during the trial: 496, 435 and 504 from the DE 110, DE 150 and the warfarin groups, respectively. DE 150 had a statistically significant reduction in reducing the risk of the stroke/SEE/PE/MI/vascular death composite endpoint when compared to warfarin (relative risk reduction of 16%, p-value 0.0096) (Table 8). DE 110 was comparable to warfarin for this endpoint.

Table 8 Hazard ratios and 95% CI for stroke/SEE/PE/MI and vascular death

	DE 110 mg vs. Warfarin	DE 150 mg vs. Warfarin
#Events/N	493/6015 vs. 496/6022	433/6076 vs. 496/6022
Hazard ratio (SE)	0.98 (0.06)	0.84 (0.05)
95% CI	0.86, 1.10	0.74, 0.96
P-value	0.6972	0.0096

[Source: reviewer's results]

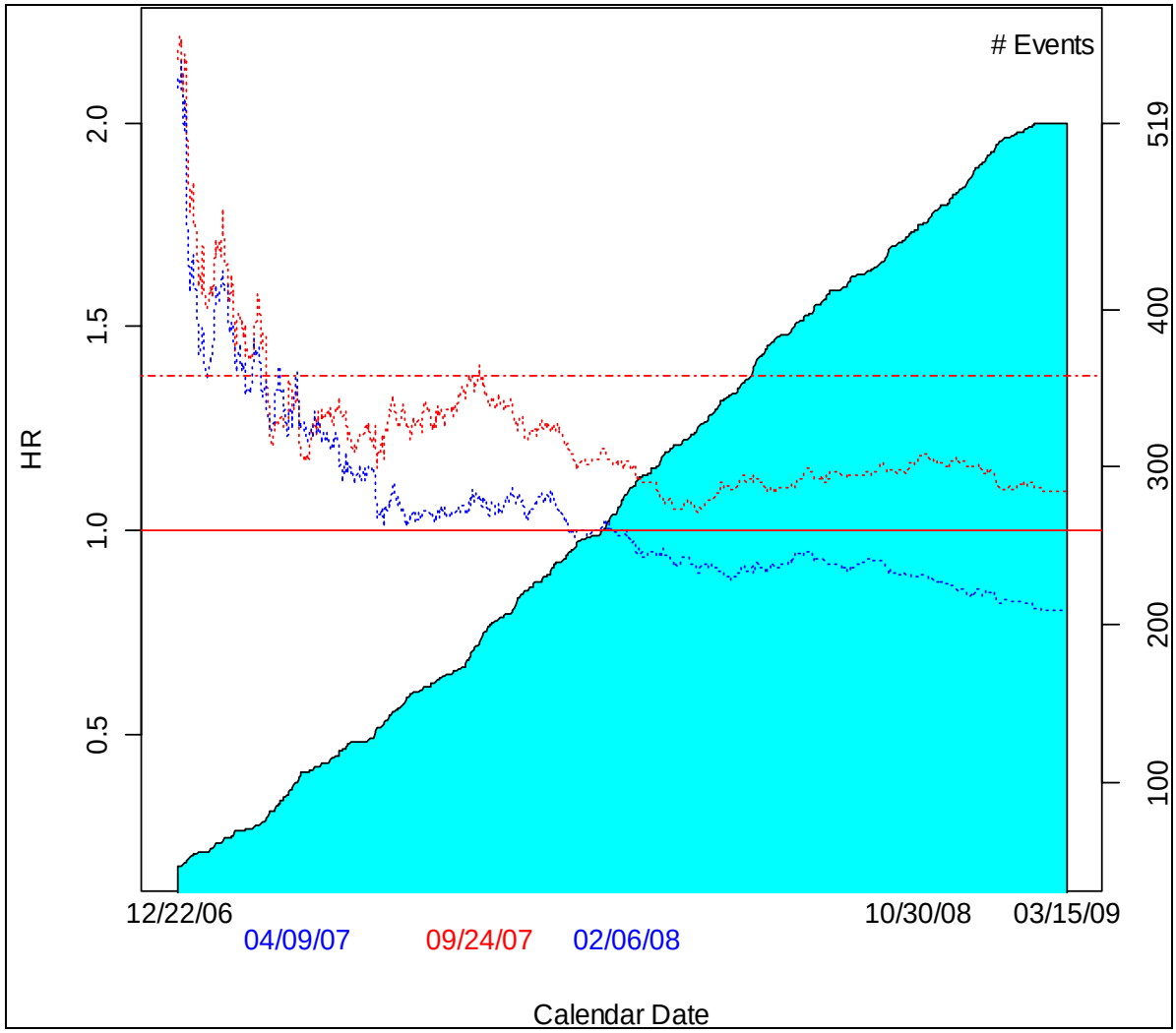
3.1.1.5 Reviewer's Results

Analysis on the Impact of Different End of Trial Dates

Both dabigatran doses achieved non-inferiority and DE 150 achieved superiority in relationship to warfarin with extremely significant statistical evidences (p-values are well less than 0.05). It would be very useful to find out how early those findings were established during the course of the trial. Figure 3.3 shows the upper 95% confidence bounds for the primary endpoint as a function of calendar time of the study. In this figure, I changed the event (censor) status and time to event information as if the current calendar time is assumed be the end of trial date starting from 12/22/2006 to 03/15/2009 (actual trial ending date). The original Cox regression analysis with treatment in the model was performed for each day to 03/15/2009. The red curve is the upper bound of hazard ratio of DE 100 mg over warfarin, and the blue curve is the upper bound of hazard ratio of DE 150 mg over warfarin. The dates on the x-axis correspond to a few important milestone dates. 12/22/2006 was arbitrarily chosen at one year after the initiation of the trial. 04/09/2007 was last time the upper bound of DE 150 mg stayed above NI margin of 1.38. 09/24/2007 was last time the upper bound of DE 100 mg stayed above NI margin of 1.38.

02/06/2008 was last time the upper bound of DE 150 mg stayed above superiority margin of 1.00. 10/30/2008 was the date that 450th adjudicated event had occurred. The blue background shows the cumulative number of events. The red dash horizontal line is the non-inferiority margin of 1.38, and the red solid horizontal line is the hazard ratio of 1.0.

Figure 3.3 The Upper Bound of Hazard Ratios for composite endpoint of stroke/SEE across trial calendar date



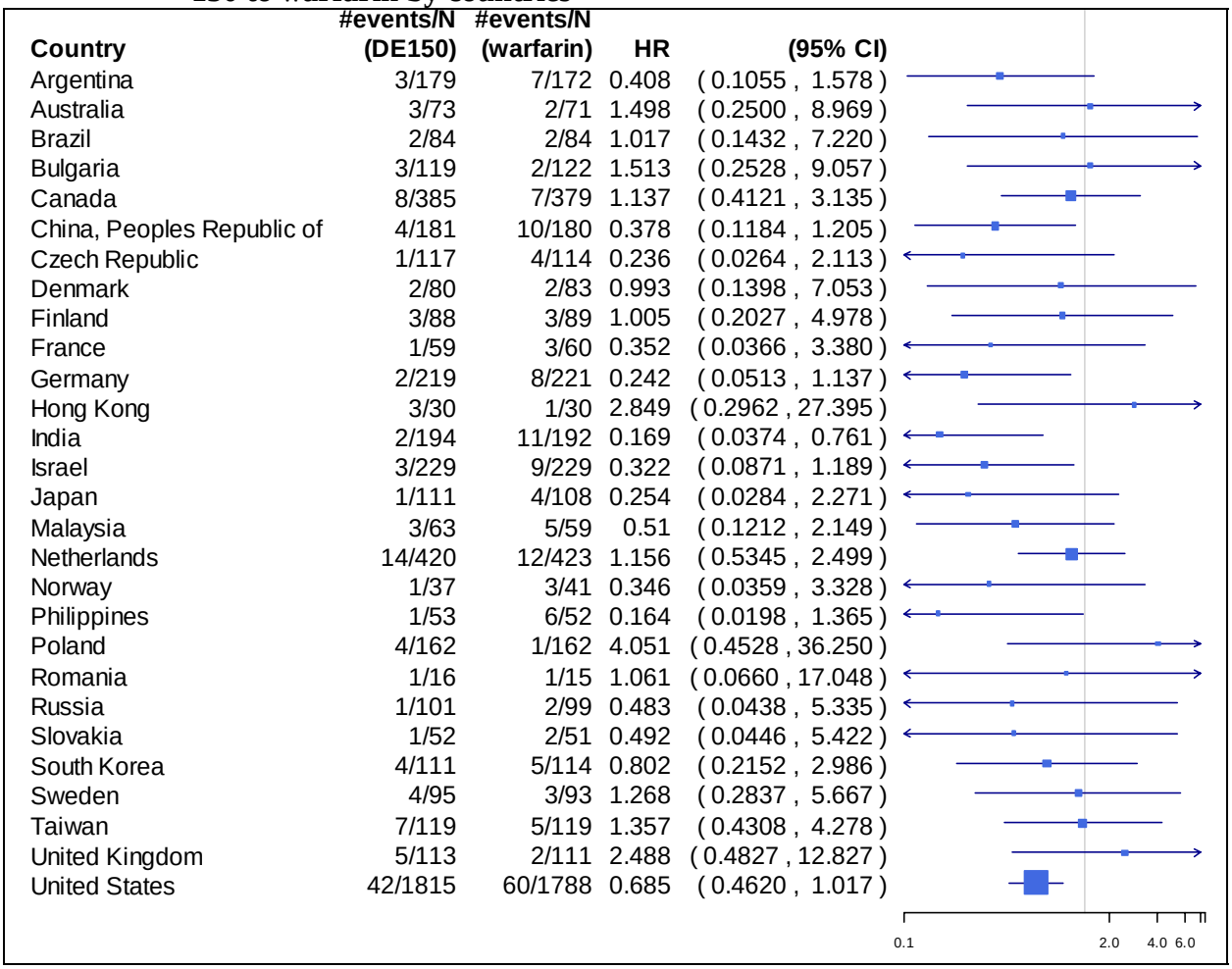
[Source: Reviewer's Results]

There are several interesting findings in Figure 3.3 should be noted. First of all, the efficacy of both dabigatran doses are very robust, the non-inferior findings were established long before the end of trial. Secondly, DE 150 mg achieved the superiority over warfarin more than one year before the end of trial. Finally, if the sponsor did not increase the originally proposed sample size from 15,000 to 18,000, the trial would still be able to demonstrate the non-inferiority claim over warfarin. When the 450th event had occurred on 10/30/2008, the study already established the overwhelming statistical evidence for the efficacy claims of the primary analysis.

Analysis on the Impact of Individual Country

The study was conducted in 44 countries. The number of patients per country ranged from 13 to 5,383. Among these countries, dabigatran doses were numerically non-inferior to warfarin in the vast majority of countries (see Figure 3.4 and Figure 3.5).

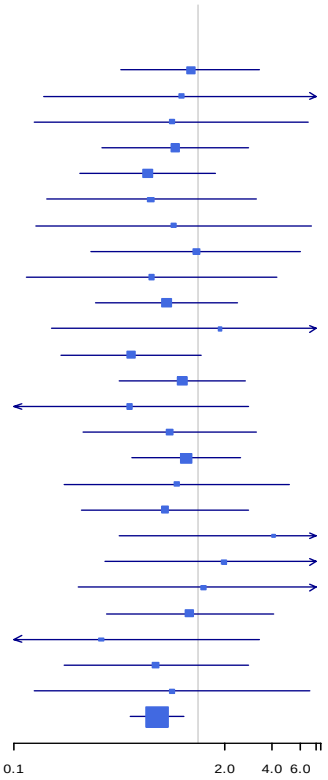
Figure 3.4 The Forest Plots of Hazard ratio and 95% CI for stroke/SEE comparing DE 150 to warfarin by countries



[Source: Reviewer's Results]

Figure 3.5 The Forest Plots of Hazard ratio and 95% CI for stroke/SEE comparing DE 110 to warfarin by countries

Country	#events/N (DE110)	#events/N (warfarin)	HR	(95% CI)
Argentina	9/179	7/172	1.242	(0.4625 , 3.34)
Australia	2/65	2/71	1.079	(0.1519 , 7.66)
Bulgaria	2/122	2/122	0.944	(0.1329 , 6.70)
Canada	7/386	7/379	0.991	(0.3475 , 2.82)
China, Peoples Republic of	7/180	10/180	0.669	(0.2545 , 1.76)
Czech Republic	3/119	4/114	0.707	(0.1582 , 3.16)
Denmark	2/84	2/83	0.97	(0.1366 , 6.88)
Finland	4/88	3/89	1.34	(0.3000 , 5.99)
France	2/58	3/60	0.706	(0.1179 , 4.22)
Germany	7/219	8/221	0.876	(0.3177 , 2.42)
Hong Kong	2/30	1/30	1.891	(0.1713 , 20.86)
India	6/192	11/192	0.532	(0.1966 , 1.44)
Israel	10/231	9/229	1.093	(0.4443 , 2.69)
Japan	2/107	4/108	0.52	(0.0952 , 2.84)
Malaysia	5/63	5/59	0.922	(0.2665 , 3.19)
Netherlands	14/423	12/423	1.163	(0.5378 , 2.51)
Norway	3/37	3/41	1.012	(0.2038 , 5.02)
Philippines	5/52	6/52	0.858	(0.2619 , 2.81)
Poland	4/161	1/162	4.048	(0.4524 , 36.22)
Russia	4/100	2/99	1.987	(0.3639 , 10.85)
Slovakia	3/53	2/51	1.5	(0.2507 , 8.98)
South Korea	6/111	5/114	1.225	(0.3739 , 4.02)
Sweden	1/89	3/93	0.344	(0.0358 , 3.30)
Taiwan	4/117	5/119	0.752	(0.2017 , 2.80)
United Kingdom	2/113	2/111	0.948	(0.1335 , 6.73)
United States	46/1780	60/1788	0.765	(0.5213 , 1.12)



[Source: Reviewer's Results]

The point estimate (hazard ratio) in the most of countries is below the non-inferiority margin of 1.38. Furthermore, the upper bounds of hazard ratio were well below the margin in United States for both dabigatran doses.

3.2 Evaluation of Safety

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

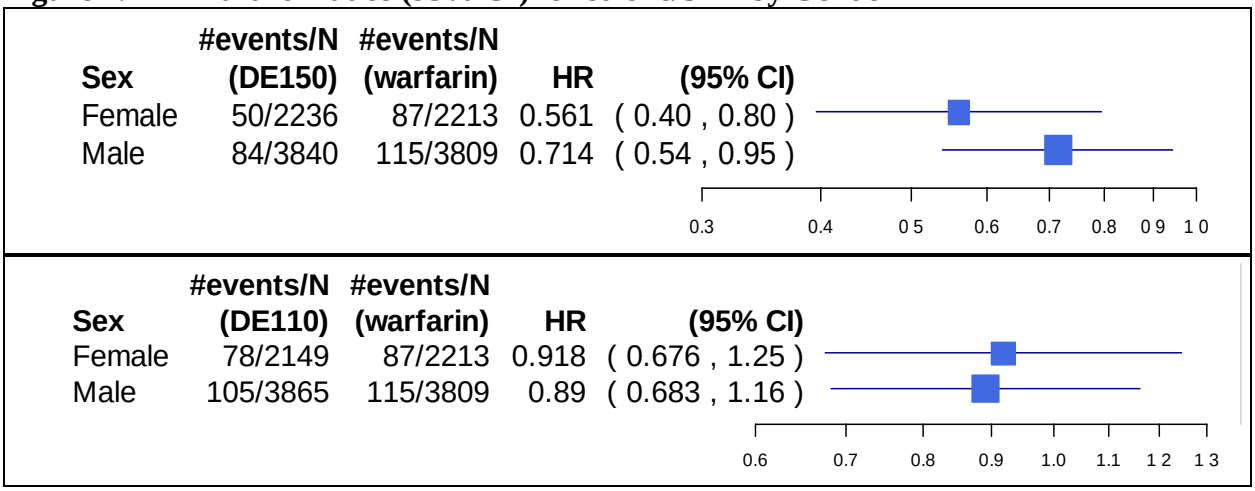
4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Age and Race group

4.1.1 GENDER

There were no obvious differences in hazard ratios for the primary endpoint across either Gender group. Both groups had favorable non-inferior results towards dabigatran doses when compared to warfarin. The DE 150mg was superior to warfarin in both female and male, see Figure 4.1.

Figure 4.1 Hazard Ratios (95% CI) for stroke/SEE by Gender

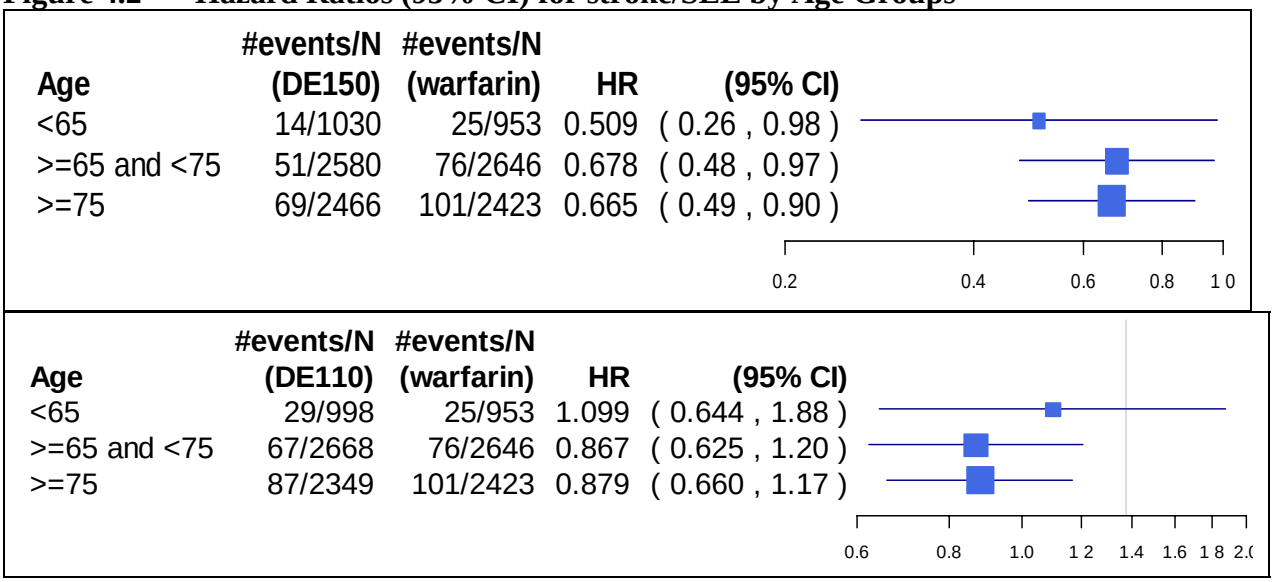


[Source: Reviewer's results]

4.1.2 AGE

The age is categorized into the following three groups: < 65, 65-75, and ≥75. The rates of stroke/SEE increased with age across all three treatment groups. Among the six comparisons in the Figure 4.2, only DE 110 had a hazard ratio greater than 1 over warfarin in the younger than 65 years of age group. The rest of groups had consistent results as the primary analysis results.

Figure 4.2 Hazard Ratios (95% CI) for stroke/SEE by Age Groups

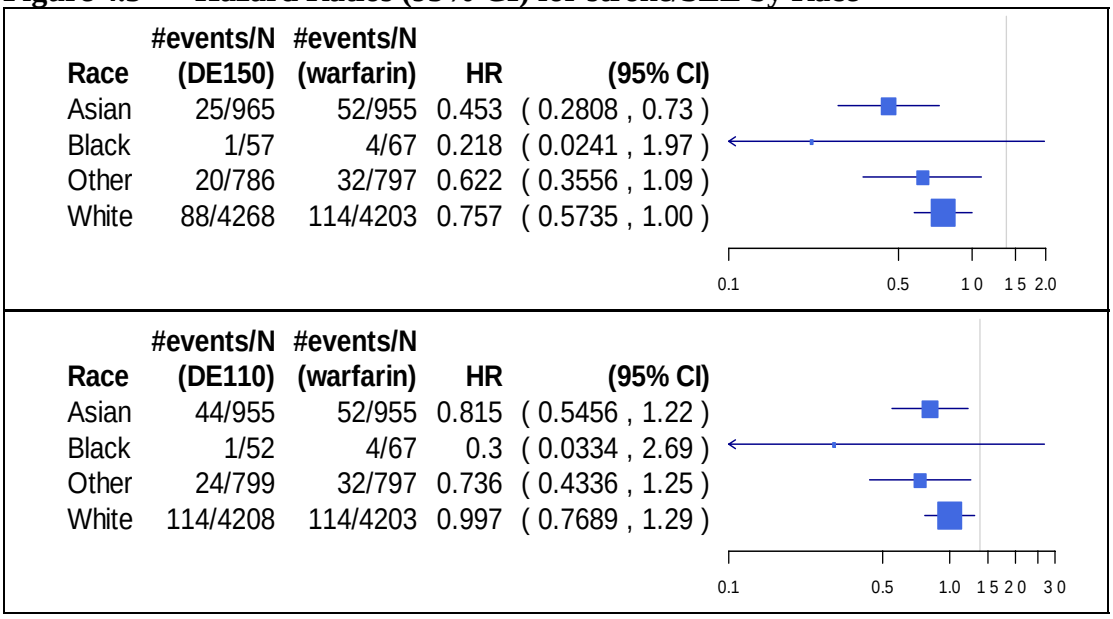


[Source: Reviewer's results]

4.1.3 RACE

Whites dominated the numbers of subjects. There were no obvious differences in hazard ratios for the primary endpoint observed across different Race groups, except that for Blacks both DE groups had relatively low hazard ratios compared to warfarin. This is due to the fact there are fewer than 70 black subjects in each group.

Figure 4.3 Hazard Ratios (95% CI) for stroke/SEE by Race



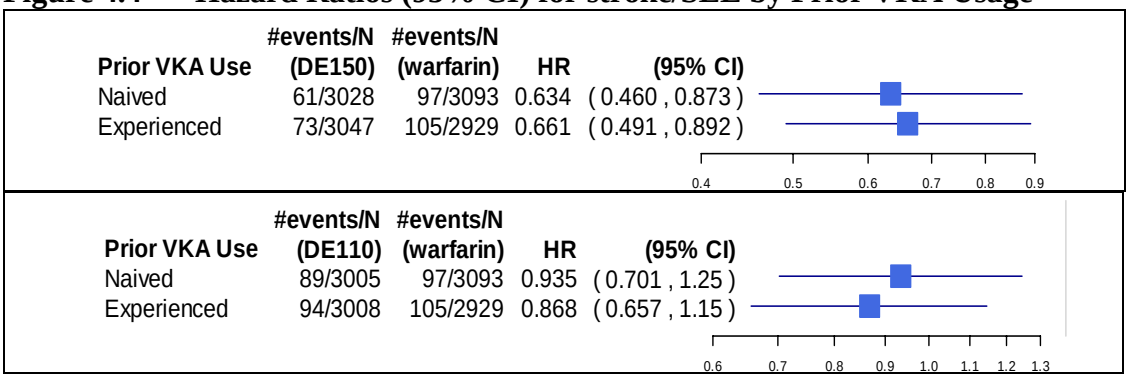
[Source: Reviewer's results]

4.2 Other Subgroup Populations

4.2.1 PRIOR VKA USE

Warfarin, the most widely used VKA, was chosen as the active control. Therefore, it is important to find out whether Dabigatran has any different effects depend on the patients' prior VKA usage.

Figure 4.4 Hazard Ratios (95% CI) for stroke/SEE by Prior VKA Usage



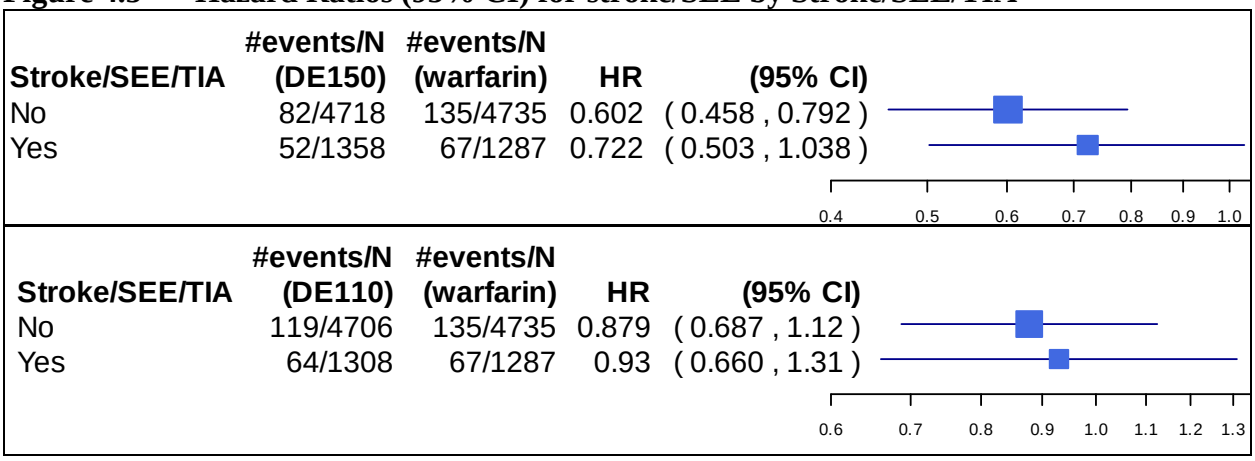
[Source: Reviewer's Results]

Based on Figure 4.4, the hazard ratio of stroke/SEE on DE 110 over warfarin did not rule out 1.00 regardless of VKA use, so they had similar event rates for this primary endpoint. On the other hand, DE 150 seemed numerically superior to warfarin regardless of VKA use.

4.2.2 HISTORY OF STROKE/SEE/TIA

The majority of subjects never had any episodes of Stroke/SEE/TIA in all treatment groups. Both Dabigatran doses had lower hazard ratio over warfarin in this subgroup. Furthermore, DE 150 was numerically superior to warfarin in this subgroup. Among the subjects who ever had history of Stroke/SEE/TIA, DE 110 seems similar to warfarin in the Stroke/SEE event rates. DE 150 nearly demonstrated superiority over warfarin in this subgroup, see Figure 4.5.

Figure 4.5 Hazard Ratios (95% CI) for stroke/SEE by Stroke/SEE/TIA



[Source: Reviewer's Results]

4.2.3 SUBGROUP ANALYSIS FOR BASELINE MEDICATION USE

The treatment effects of both DE 110 and 150 were generally consistent across all subgroups defined by baseline medication use in comparison to warfarin. In general, DE 110 was comparable to warfarin and DE 150 was numerically superior to warfarin for most of the subgroups, see Table 9.

Table 9 Hazard Ratios (95% CI) for Stroke/SEE by medication use

	DE 110 mg vs. Warfarin	**DE 150 mg vs. Warfarin**
	Hazard Ratio (95% CI)	Hazard Ratio (95% CI)
ASA		
Never used	0.92 (0.71, 1.20)	0.64 (0.48, 0.86)
Used at least once	0.87 (0.64, 1.18)	0.67 (0.48, 0.93)
Clopidogrel		
Never used	0.95 (0.77, 1.17)	0.66 (0.52, 0.83)
Used at least once	0.38 (0.16, 0.91)	0.57 (0.27, 1.20)
ASA+Clopidogrel		
Never used	0.96 (0.78, 1.17)	0.67 (0.48, 0.86)
Used at least once	0.08 (0.01, 0.58)	0.31 (0.10, 0.93)
Amiodarone		
Never used	0.94 (0.77, 1.16)	0.68 (0.54, 0.85)
Used at least once	0.47 (0.21, 1.04)	0.35 (0.15, 0.83)
Verapamil		
Never used	0.94 (0.76, 1.16)	0.67 (0.54, 0.84)

Used at least once	0.46 (0.20, 1.06)	0.40 (0.17, 0.96)
Diltiazem		
Never used	0.93 (0.76, 1.15)	0.67 (0.53, 0.84)
Used at least once	0.60 (0.29, 1.23)	0.46 (0.21, 1.02)
Statin		
Never used	1.00 (0.77, 1.31)	0.56 (0.41, 0.76)
Used at least once	0.78 (0.58, 1.06)	0.75 (0.55, 1.02)

[Source: Reviewer's Results]

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The non-inferiority margin

The proposed testing hypothesis for RE-LY was that whether either dabigatran doses (110 mg and 150 mg) were non-inferior to warfarin in reducing the incidences of Stroke/SEE. The non-inferiority margin of 1.46 for the hazard ratio in the sponsor's study report was derived based on the historical placebo controlled trials using the 95%-95% rule. This rule utilized the lower bound of the 95% confidence interval of the hazard ratio for warfarin versus placebo for the derivation of the non-inferiority margin, and the upper bound of the 95% confidence limit for dabigatran versus warfarin for the statistical test. The margin 1.46 used in the study design preserved at least 50% of warfarin's effect on the risk ratio scale using the lower bound of the 95% of the risk ratio of placebo over warfarin. However, a smaller margin of 1.38, derived to preserve the effect of warfarin on the Log scale, was recommended by a regulatory agency. In spite of this discrepancy on the margin, both dabigatran doses were non-inferior to warfarin based on the sponsor's efficacy findings.

Summary of the historical trials and constancy assumption

The effectiveness of warfarin has been studied both in placebo-controlled and active-controlled trials. There are six placebo-controlled studies of warfarin involved the patients with AF between 1989 and 1992. All these trials showed a consistent efficacy for warfarin in preventing stroke and other cardiovascular events, despite differences in their designs and patient populations. The primary outcomes of these studies are summarized in Table 10. Almost all of the trials showed significant reduction in the primary endpoint event by warfarin against placebo. Trial CAFA failed to show a significant benefit over placebo, but the estimated warfarin effect from this trial was consistent with those observed from the other trials.

Table 10 Placebo-Controlled Trials of Warfarin in Non-Valvular Atrial Fibrillation

Study	Summary	Events/Patient Years		Risk Ratio (95% CI)
		Warfarin	Placebo	
AFASAK	open label. 1.2 yr follow-up	9/413 = 2.18%	21/398 = 5.28%	0.41 (0.19, 0.89)
BAATAF	open label. 2.2 yr follow-up	3/487 = 0.62%	13/435 = 2.99%	0.21 (0.06, 0.72)
EAFT	open label. 2.3 yr follow-up patients with recent TIA	21/507 = 4.14%	54/405 = 13.3%	0.31 (0.19, 0.51)
CAFA	double blind. 1.3 yr follow-up	7/237 = 2.95%	11/241 = 4.56%	0.65 (0.26, 1.64)
SPAF I	open label. 1.3 yr follow-up	8/260 = 3.08%	20/244 = 8.20%	0.38 (0.17, 0.84)
SPINAF	double blind. 1.7 yr follow-up	9/489 = 1.84%	24/483 = 4.97%	0.37 (0.17, 0.79)

[Source: FDA Non-inferiority draft guidance Table 1]

Even if the historical studies are consistent, a critical consideration in deciding upon the NI margin derived from these studies is whether the constancy assumption is reasonable. To evaluate the plausibility of this constancy assumption, one might compare some features of the six placebo-controlled warfarin studies with the RE-LY study. There is considerable heterogeneity in the demographic characteristics of these studies. The draft guidance listed number of characteristics, such as a history of stroke or TIA, see Table 11. The most of characteristics are similar among the historical studies with RE-LY, but the history of stroke or TIA and CAD are much higher in RE-LY.

Table 11 Comparisons on Demographic Variables, Clinical Characteristics, and Endpoints of Historical Warfarin AF Studies vs. RE-LY

	AFASAK	BAATAF	CAFA	SPAF	VA	EAFT	RE-LY
Age years (mean)	73	69	68	65	67	71	71
Sex (%) Male	53%	75%	76%	74%	100%	59%	59%
h/o stroke or TIA (%)	6%	3%	3%	8%	0%	100%	20%
h/o HTN (%)	32%	51%	43%	49%	55%	43%	NA
≥65 years old & CAD (%)	8%	10-16%	12-15%	7%	17%	7%	24.2%
>65 years old & DM (%)*	7-10%	14–16%	10-14%	13%	17%	12%	19.3%
h/o LV dysfunction (%)*	50%	24-28%	20-23%	9%	31%	8%	10.7%
Mean BP at BL (mm Hg)	NA	NA	NA	130/78	NA	145/84	131/77
Target INR	2.8-4.2	1.5-2.7	2-3	2-4.5	1.4-2.8	2.5-4.0	2-3
Primary endpoint	Stroke, TIA, systemic embolism	Ischemic stroke	Ischemic stroke and systemic embolism	Ischemic stroke and systemic embolism	Ischemic stroke	Vascular death, NF MI, stroke, systemic embolism	Stroke and SEE

Increase of sample size

The study was originally designed as an event driven trial. Based on an estimated yearly event rate of 1.6% and a two-year enrolment period and one-year follow up, a total of 15,000 subjects were planned to be randomized from approximately 800 centers. Due to rapid enrollment, 15,000 subjects were randomized in 1.5 years (18 months). Sponsor claimed that if the recruitment was stopped at that time, the last randomized subjects would have had to follow up for more than 1 year to achieve the planned total number of events, if the actual event rate was as expected. In addition, sponsor also claims that the actual event rate could be less than 1.6% based on other published studies. Therefore, sponsor decided to continue the recruitment as planned, which resulted a total of 18,113 subjects were randomized and the total number of subjects with adjudicated stroke/SEE was increased to 519. The above changes were added to protocol's second amendment on May of 2007. In order to validate the final primary efficacy results, both the sponsor and this reviewer had performed the sensitivity analysis for the first 450 adjudicated primary events. Based on all the analyses results, both doses of dabigatran have met the pre-

specified non-inferior margin to warfarin to conclude that the two doses are effective for the stroke prevention in AF patients. Furthermore, DE 150 was superior to warfarin as well, see Table 4.

5.2 Conclusions and Recommendations

Overall, RE-LY demonstrated that both doses of dabigatran were non-inferior to warfarin and DE 150 was superior to warfarin for the primary (stroke/SEE) efficacy endpoints. Furthermore, the secondary (stroke/SEE/death and stroke/SEE/PE/MI/vascular death) efficacy endpoints also met the above claims numerically. However, sponsor did not specify the statistical testing rules and margins for these endpoints in the TSAP. Therefore, these findings can only be viewed as exploratory findings.

There was no discrepancy results found in any of the sensitivity analyses. Although, DE 150 did not show superiority for US subjects statistically, but it was still non-inferior to warfarin and the point estimate (hazard ratio) also less than 1.00. All the subgroup analyses performed in Section 4 were consistent with the primary efficacy results. Hence, RE-LY's finding is very robust. Furthermore, based on the reviewer's analysis on the impact of different end of trial dates, the dabigatran doses achieved the non-inferiority long before the end of trial date and DE 150 achieved superiority to warfarin more than one year before the end of trial date, see Figure 3.3.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22512	ORIG-1	BOEHRINGER INGELHEIM PHARMACEUTICALS INC	PRADAXA (DABIGATRAN ETEXILATE MESYLATE)

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
07/20/2010

HSIEN MING J J HUNG
07/20/2010

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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	22,512
Drug Name:	Pradaxa™, Dabigatran Etexilate Mesylate Capsules
Indication:	Stroke Prevention in Atrial Fibrilation
Applicant:	Boehringer Ingelheim Pharmaceuticals, Inc. Ridgefield, Connecticut
Date:	Submitted 29 March 2009
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Cardiovascular and Renal Products
Toxicologist:	Reviewer: Patricia Harlow, Ph.D. Albert Defelice, Ph.D.
Project Manager:	Allison Blaus
Keywords:	Bayesian analysis, Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY

According to the Sponsor the objective of this study was to assess the carcinogenic potential of compound BIBR 1048 MS, a thrombin inhibitor with proposed name "Pradaxa", when administered orally by gavage to HsdBrl Han:Wist (Han Wistar) rats and CD-1 mice for 104 weeks. Both studies were conducted at the (b) (4)

1.1. Conclusions and Recommendations

In both the rat and mice studies there were four treatment groups, including the vehicle control group. The Sponsor reports that the oral route of administration was chosen to simulate the conditions of clinical administration. In both genders in rats, there were three dose groups, each with 55 Han Wistar Rats, dosed by either the vehicle, 0.5% Natrosol 250 HX solution, or BIBR 1048 MS/Pradaxa at dosages of 30 or 100 mg/kg/day. A larger fourth dose group with 65 rats in each gender received the test drug at a dosage of 200 mg/kg/day. Similarly in both genders in mice, there were three dose groups, each with 54 CD-1 mice, dosed with either the vehicle above, and the test drug also at dosages of 30 or 100 mg/kg/day, plus a fourth treatment group with 63 mice per gender dosed at 200 mg/kg/day dose. In each study, these dose groups, groups 1-4, are also labeled as "Vehicle", "Low", "Medium", and "High."

Simple summary life tables in mortality are presented in Tables 10, 11, 15, and 16 below. Kaplan-Meier estimated survival curves across dose groups for each gender in each species are displayed in Figures A.1.1-A.1.4 in Appendix 1. The statistical significances of the tests of differences in survival across treatment groups are given in Table 1 below. Tests of homogeneity over all groups, dose related trend and pairwise differences between the high dose group and the vehicle controls were performed.

Table 1. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Rats				
Homogeneity over all 4 Groups	0.0400	0.0477	0.0004	<0.0001
Trend over Groups Vehicle to High	0.0070	0.0096	<0.0001	<0.0001
Comparison of Vehicle and High	0.0060	0.0072	<0.0001	<0.0001
Mice				
Homogeneity over all 4 Groups	0.9519	0.8547	0.3065	0.3258
Trend over Groups Vehicle to High	0.8735	0.9560	0.0950	0.0814
Comparison of Vehicle and High	0.9227	0.6554	0.1563	0.1202

From Table 1. above, there is evidence of differences in survival curves in rats, particularly female rats (Male rat Logrank $p = 0.0400$, Wilcoxon $p = 0.0477$, Female rat Logrank $p = 0.0004$, Wilcoxon $p < 0.0001$). From the Kaplan-Meier product limit estimated survival curves in Figures A.1.1 and A.1.2 below one can see that to some degree in male rats, and especially in female rats, there is evidence of a general increase

in mortality with increasing dose. This consistent with the overall tests of homogeneity cited above above and even more with the more powerful test of trend over dose (Male rats trend: Logrank p = 0.0070, Wilcoxon p = 0.0072, Female rats trend: both p<0.0001). Results for the pairwise comparison between the vehicle and high dose group were similar to the results for trend. In mice there was no strong evidence of heterogeneity in survival over dose. From figure A.1.3 in male mice there appears to be suggestion of a roughly increasing mortality with decreasing dose, although survival curves are closely intertwined at the end of the study. Figure A.1.4 suggests a slight increase in mortality over increasing dose in female mice, but, nonetheless survival curves do seem to be closely intertwined. In particular, in male mice the vehicle dose group has generally the highest or close to highest mortality, with the medium dose group next, but closely intertwined with the high dose group. Finally the low dose group seems to roughly have the lowest mortality. However, no tests of lack of overall homogeneity, pairwise homogeneity, or trend were statistically significant at the usual 0.05 level in either mouse gender, (Males: all p ≥ 0.6554, Females: all p ≥ 0.0814).

Appendix 2 presents an experimental Bayesian analysis of dose related survival based on proportional hazards models. It should be noted that unlike the usual formulation for proportional hazards, this analysis allows intersecting survival curves.

Table 2, below, lists tumors and the p-values using the so-called poly-k modification of the Cochran-Armitage test for all tumors having any trend or pairwise comparison to vehicle with statistical significance less than 0.05, and thus are potentially statistically significant (please see Section 1.3.1.3). Complete incidence tumor tables are given in Appendix 3. Applying the Haseman-Lin-Rahman rules (please see Section 1.3.1.4) to adjust for the multiplicity of tests being performed, the test for trend in rare tumors (incidence ≤ 1%) could be considered statistically significant if the observed p-value is 0.025 or less, while common tumors (incidence > 1%) would be considered as statistically significant if the observed p-value is 0.005 or less. The pairwise test between the high dose group and the vehicle controls could be considered statistically significant if the observed p-value is 0.05 or less in rare tumors, and 0.01 or less in common tumors. Note that applying these rules to the specific tests for pairwise comparisons between the vehicle controls and the low and medium dose groups alone would be expected to increase the overall type I error rate to above the nominal rough 10% level.

Table 2. Potentially Statistically Significant Neoplasms

Organ Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
Female Rats								
OVARIES								
B-GRANULOSA CELL TUMOUR	0	0	2	4	0.0038	0.0326	0.1707	.
Sertoli/Granulosa Cell Tumor	1	0	2	4	0.0142	0.1083	0.3747	0.4891
Male Mice								
H'POIETIC TUMOUR								
PLEOMORPHIC LYMPHOMA	0	3	5	2	0.3035	0.2636	0.0217	0.1249
SKIN (PROTOCOL)								
Fibroma/Fibrosarcoma/Sarcoma	0	4	3	6	0.0418	0.0182	0.0948	0.0555

Using the incidence in the vehicle control group to define the rarity of the neoplasms, in rats, only the test of trend in benign granulosa cell tumors were statistically significant (i.e., $p = 0.0038 < 0.025$), plus the pairwise test comparing the high dose group to control (i.e., $p = 0.0326 < 0.05$). The test of trend in pooled Sertoli and granulosa cell tumor was nominally statistically significant (i.e., $p = 0.0142$). If we ignore the fact this is a comparison to the medium dose group, since the single incidence of Sertoli cell tumor in the control group would be sufficient to classify the pooled tumor as common, and this $p > 0.005$, then this tumor would not be categorized as statistically significant. No other tests in rats even achieved the usual 0.05 level of significance.

In male mice the pairwise comparison of the medium dose group to the control was also statistically significant (i.e., $p = 0.0217 < 0.05$), however, there was no other evidence of a dose related trend (i.e., trend $p = 0.3035$ and both other pairwise comparisons have $p \geq 0.1249$). In male mice the pairwise comparison of the high dose group to vehicle in pooled fibroma, fibrosarcoma, and sarcoma of the skin was statistically significant (i.e., $p = 0.0182 < 0.05$). However, although close to significance, using the Haseman-Lin-Rahman rules to adjust for multiplicity, no other test was statistically significant (i.e., trend $p = 0.0418 > 0.025$ and both other pairwise comparisons have $p = 0.0555, 0.0948 > 0.05$). Note that no other pairwise tests or tests of trend in mice achieved statistical significance using the nominal 0.05 level.

Complete incidence tables with the results of statistical tests of trend and pairwise comparisons to vehicle are given in Appendix 3.

1.2. Brief Overview of the Studies

This submission had two studies, one with rats, the other with mice:

Study BIBR 1048 MS Carcinogenicity Study by Oral Gavage Administration to Han Wistar Rats for 104 Weeks,

and,

Study BOI287/042668 Carcinogenicity study by oral gavage administration to CD-1 mice.

Both the rat and mice studies were designed with four different treatment groups, the vehicle, 0.5% Natrosol 250 HX solution, and BIBR 1048 MS at dosages of 30, 100, or 200 mg/kg/day. In rats these four dose groups were assigned 55, 55, 55, and 65 animals, respectively, per gender. In mice the similar doses were 54, 54, 54, and 63 animals per gender. In each study, these dose groups, groups 1-4, are also labeled as "Vehicle", "Low", "Medium", and "High."

Animals were dosed once daily by oral gavage, 7 days a week for up to 24 months before necropsy.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

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

1.3.1.1. Survival Analysis:

Two main test statistics are provided, the log rank test and the so-called Wilcoxon test. The log rank tests tends to puts higher weight on later events, while the Wilcoxon test tends to weight events more equally, and thus is more sensitive to earlier differences in survival. The log rank test is most powerful when the survival curves track each other, and thus the proportional hazard assumption seems to be true. The number of such tests raises issues of multiple testing, but from the point of view of finding differences among treatment groups (i.e., reducing the probability of Type II error), this should be acceptable. Appendix 1 reviews the animal survival analyses in some detail. The Sponsor's analyses are summarized in Section 3.2.1.1. Appendix 2 presents an experimental Bayesian approach that allows nonproportional hazards.

1.3.1.2. Tests on Neoplasms:

The Sponsor based conclusions on so-called Peto analyses. These require assessment of whether or not a tumor can be classified as fatal, a classification which can be difficult to accurately determine. Further, Peto analysis of observable tumors is also based on time of detection. Other tumors are defined as "incidental" and the incidence of such tumors are analyzed using tests stratified on time period. Until recently, most submissions to CDER, were analyzed with such Peto tests. Finally, with fixed time intervals there can be computational problems with the Peto tests.

The Society of Toxicological Pathology had a town hall meeting in June 2001 where this approach was criticized. The primary alternative discussed in the commentary on this meeting (STP Peto Working Group, 2002) is the poly-k modification of the Cochran-Armitage test of trend for tumor incidence. This is the method of analysis used in the FDA analysis in this report.

1.3.1.3. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involved a large number of statistical tests, which in turn necessitated an adjustment in experiment-wise Type I error. One, perhaps the usual, approach for two species, two gender, two year studies with testing for trend over three doses and vehicle controls and comparing the high dose group to controls follows the Haseman-Lin-Rahman rules for a Peto analysis. Based on his extensive experience with such analyses, for pairwise tests between the high dose group and controls in two species, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. For a standard chronic study in two species (i.e., mice and mice) study, based on simulations and their experience, 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. This is the adjustment used by the Sponsor.

This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation). Rahman and Lin (2008) showed that these rules are also roughly applicable to the poly-k analyses used here. Note that strictly speaking, these rules only control the overall errors of the tests of trend over four doses for two genders, each with one marginal trend test and a corresponding test comparing the high dose to the control. Applying these rules to other comparisons to controls irrespective of the tests of trend and the pairwise test comparing the high dose to controls inflates the experiment-wise type I error rate, though the level of inflation is not clear. In this analysis we will use the observed incidence in the vehicle control group to decide if a tumor is rare or common.

1.3.1.4. Housing of Animals:

Rats were initially housed five animals of the same gender together in each cage, until reduced by mortality or the need to separate animals within the cage. Mice were started with three per cage, except that due to problems associated with fighting, from Week 72 onwards male animals were housed individually. Note that multiple housing of animals may cause statistical problems in the analysis. It is possible that proximity may induce positive correlations in response, while within cage competion could induce either negative or positive correlations. Because of this housing the within treatment estimated variances may be too large or too small, resulting in conservative or liberal tests (in terms of Type I error). Unless it has been clearly shown that tumor incidence is independent of cage, from a purely statistical point of view, this reviewer would generally recommend single housing of animals. Without data on the actual caging these effects can not be investigated. However, dose administration by gavage, as say opposed to dosing in the diet, would be expected to reduce the effects of these associations.

1.3.1.5. Validity of the Designs:

When determining the validity of designs there are two key points:
1) adequate drug exposure,
2) tumor challenge to the tested animals.

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

Lin and Ali (1994), quoting work by Haseman, have suggested that a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. Note this simple criterion does seem to be satisfied.

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' " The values in the following two tables were taken from the Sponsor's tables. Note that the values for the weight decrement of the high dose group in each gender in each species seem to agree with the criterion cited above. Although this is a decision for the toxicologist, this may be evidence that the MTD was met in each study.

Table 3. Relative Group Mean Weight Change (g – compared to control) in Rats

Rats Dose Label Nominal Dose mg/kg/day	Males		Females	
	Change	% from Control	Change	% from Control
Vehicle - 0	397.4		211.8	
Low - 30	385.5	-4.1%	208.4	-4.1%
Medium - 100	394.7	-1.8%	214.3	-1.4%
High - 200	348.3	-13.3%	196.3	-9.7%

Table 4. Relative Group Mean Weight Change (g – compared to control) in Mice

Mice Dose Label Nominal Dose mg/kg/day	Males		Females	
	Change	% from Control	Change	% from Control
Vehicle - 0	9.9		10.8	
Low - 30	9.5	-2.6%	11.3	-3.8%
Medium - 100	10.4	+6.7%	11.1	-5.5%
High - 200	9.1	-6.7%	10.9	-7.2%

Tables 5 and 6 below summarize overall food consumption. For rats the values are transcribed from pages 77 and 80 of Table 5 in the Sponsor's rat report.

Table 5. Mean Food Consumption in Rats (g/animal/week)[1]

Rats Dose Label Nominal Dose mg/kg/day	Males		Females	
	Overall Mean	% from Control	Overall Mean	% from Control
Vehicle - 0	144		112	
Low - 30	141	101%	115	103%
Medium - 100	144	100%	116	104%
High - 200	142	99%	117	104%

[1]Mean to Week 104

For mice the values are transcribed from page 84 of Table 5 in the Sponsor's mice report.

Table 6. Mean Food Consumption in Rats (g/animal/week)

Mice Dose Label Nominal Dose mg/kg/day	Males		Females	
	Overall Mean[1]	% from Control	Overall Mean[2]	% from Control
Vehicle - 0	39		36	
Low - 30	41	105%	37	101%
Medium - 100	40	103%	36	100%
High - 200	39	100%	36	100%

[1]Mean to Week 104
[2]Mean to Week 100

Thus, at least from these values, neither species seems to show any clear dose related trend in food consumption.

Again from 2) in the leading paragraph of this section, excess mortality, not associated with any tumor or sacrifice in the higher dose groups, might suggest that the MTD was exceeded. One way to assess this possibility is to measure mortality not associated with any identified tumor. This seems to be a new way to assess if the high dose is at the MTD. Tables 7 and 8 below indicate the number of animals in each dose group in each study that died of a natural death or moribund sacrifice, but did not show any tumors:

Table 7. Natural Death with No Identified Tumor in Rats

Group Label	Dose Mg/kg/day	Males		Females	
		Died w/o tumor	Other	Died w/o tumor	Other
Vehicle	0	1 2%	54 98%	1 2%	54 98%
Low	30	6 11%	49 89%	2 4%	49 96%
Medium	100	12 22%	43 78%	14 25%	43 75%
High	200	17 26%	48 74%	19 29%	48 71%

To compare the incidence of deaths without tumors we can specify the usual survival tests where those animals that die with a tumor or are sacrificed are considered as censored. The remaining animals are those that die prior to developing a tumor. If the MTD is achieved we would expect a slight increase in animals that die before developing a tumor. If the MTD is exceeded we would expect a larger dose related excess toxicity, resulting in a dose related trend in these deaths. First, note that the results on survival in rats indicate generally increasing mortality in increasing dose. In the table above, the fact the high dose group is larger than the others suggest the percentage of animals in the dose group that die without tumor is more interpretable than the simple counts. However in both rat genders in the tables above there is clear evidence of what might be termed dose related "premature deaths." This is confimed by the tests comparing the event

history curves of the high dose to the vehicle (Males: Logrank p=0.0002, Wilcoxon p=0.0003, and Females: both Logrank and Wilcoxon p < 0.0001).

Table 8. Natural Death with No Identified Tumor in Mice

Group Label	Dose Mg/kg/day	Males		Females	
		Died w/o tumor	Other	Died w/o tumor	Other
Vehicle	0	18 33%	36 67%	9 17%	45 83%
Low	30	13 24%	41 76%	2 4%	52 96%
Medium	100	18 33%	36 67%	13 24%	41 76%
High	200	20 32%	43 68%	17 27%	46 73%

In mice neither the tables above, nor the tests comparing the event history curves of the high dose to the vehicle (Males: Logrank p=0.1436, Wilcoxon p=0.1006, and Females: Logrank p=0.7258, Wilcoxon p=0.6040) show any strong evidence of such early deaths. It should be emphasized that these are all merely observations. Determination of the MTD status requires the expertise of the toxicologist.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

According to the Sponsor the objective of this study was to assess the carcinogenic potential of Pradaxa, compound BIBR 1048 MS, a thrombin inhibitor, when administered by gavage to Han Wistar rats and CD-1 mice for 104 weeks. Both studies were conducted at (b) (4)

2.2. Data Sources

SAS transport data sets for each species were provided by the Sponsor and placed in the edr. In rats the tumor data set was tumor288.sas7bdat with a summary of survival in dthtime.sas7bdat. In mice the tumor data set was tumxpt4.sas7bdat with a summary of survival in dthtime.sas7bdat.

In the tumor data sets, for each species, time of death or sacrifice was provided in both weeks and days. However, for mortality independent tumors the time of detection was provided in weeks only. It was decided to analyze the time of tumor detection in units of days, estimated by the lowest of the time of detection and the time of death. But, again the former were measured only in weeks. To keep the time units in days the time

to tumor detection was estimated as death time in days – 7 times the difference of the time to death in weeks minus the week of detection. This means the nominal time of detection may differ slightly from the true value, but should be within the same week, with a smaller error than would be obtained by using just the by week values.

Survival analyses were based on actual time of death or sacrifice.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

More detailed results on the study are presented below.

3.2.1. Study BIBR 1048 MS Carcinogenicity Study by Oral Gavage Administration to Han Wistar Rats for 104 Weeks.

STUDY DURATION: 105/106 Weeks.
DOSING START DATE: June 19, 2003
TERMINAL SACRIFICE (NECROPSY) DATES: June 20 – June 28, 2005.
STUDY ENDING DATE (Completion of experimental work) June 27, 2005.
MOUSE STRAIN: Rat (Crl:CD[SD]IGSBR VAF/Plus).
ROUTE: Gavage

The carcinogenic potential of Pradaxa, i.e. compound BIBR 1048 MS (a thrombin inhibitor), to HsdBrl Han:Wist (Han Wistar) rats by oral administration was assessed over a period of 104 weeks. Two groups, each comprising 55 male and 55 female rats received BIBR 1048 MS at dosages of 30 or 100 mg/kg/day. A further group, comprising 65 male and 65 female rats received BIBR 1048 MS at a dosage of 200 mg/kg/day. A Control group of 55 male and 55 female rats received the vehicle, 0.5% Natrosol 250 HX solution, at the same volume-dosage. A further five males and five females were allocated to the control group treatment, and a further 10 males and 10 females in the remaining dose groups were used for toxicokinetic analyses.

The Sponsor reports that the "dosages used in this study (0, 30, 100 and 200 mg/kg/day) were selected in conjunction with the Sponsor on the basis of data available from a 2-week dose-range finding study (BOI 266/020214) and a 13-week study (BOI 277/032919) in the Han Wistar rat. In the 13-week study, 300 mg/kg/day was associated with mortality due to haemorrhage, an expected consequence of the pharmacology of BIBR 1048 MS (a thrombin inhibitor), precluding the administration of this dosage for a long-term study. The Low dosage of 30 mg/kg/day was intended to provide information on the NOTEL. It was estimated that this dosage would result in plasma concentrations

of BIBR 1048 MS in excess (>2-fold) of those expected in maximum in man at the clinical trial dosages employed to date (up to 300 mg per patient). A larger group size was used for the highest dosage group on this study to allow for the potentially higher mortality due to the pharmacological effects of BIBR 1048 MS." (pages 11-12 of report)

The Sponsor reports that five animals of the same gender were housed together in each cage, until reduced by mortality or the need to separate animals within the cage. Food and water were available ad libitum, except when urine was being collected.

3.2.1.1. Sponsor's Results and Conclusions

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

Survival analysis:

The Sponsor summarizes the results of the log rank tests comparing survival curves in rats as follows:

"Males
The trend test, when all treated groups were included, showed a statistically significant increase in mortality with increasing dose ($p=0.007$). When the top dose group was excluded, the trend test still showed a significant increase in mortality ($p=0.018$). The pairwise comparisons showed that the 100 and 200 mg/kg/day groups had significantly higher mortality than the control group ($p=0.038$ and $p=0.009$ respectively).

"Females
The trend test, when all treated groups were included, showed a statistically significant increase in mortality with increasing dose ($p<0.001$). When the top dose group was excluded, the trend test still showed a significant increase in mortality ($p=0.013$). The pairwise comparisons showed that the 100 and 200 mg/kg/day groups had significantly higher mortality than the control group ($p=0.035$ and $p<0.001$ respectively)." (page 288 of report).

Tumorigenicity analysis:

The Sponsor reported the results of a Peto analysis using the Haseman-Lin-Rahman rules in Table 9 below, transcribed from page 300 of the Sponsor's report.

Table 9. Interpretation of p-values using Common/Rare classification

Sex	Tissue	Tumour	Historical Incidence (1)	Class	Statistical comparison and p-value		Threshold (2)	Significant
Male	Testes	Benign interstitial (Leydig) cell adenoma	2.5% (12/477)	Common	Trend test including all groups	0.020	0.005	No
					Pairwise comparison of Control vs Group 4	0.032	0.01	No
Female	Ovaries	Benign granulosa cell tumour	3.4% (16/476)	Common	Trend test including all groups	0.011	0.005	No

(1) Based on control animals from nine similar studies.
(2) Based on FDA Draft Guidance.

The Sponsor's reported tumor incidence seems to generally agree with the results in Appendix 3. The results of the Sponsor's analysis are summarized verbally as follows:

"Males
Testes
For benign interstitial (Leydig) cell adenoma, the trend test was statistically significant when all the groups were included in the analysis (p=0.020). Upon exclusion of Group 4, the trend test was no longer statistically significant (p=0.280). The pairwise comparison between the control group and Group 4 was statistically significant (p=0.032).
Females
Ovaries
For benign granulosa cell tumour, the trend test was statistically significant when all the groups were included in the analysis (p=0.011). Upon exclusion of Group 4, the trend test was no longer statistically significant (p=0.078). None of the pairwise comparisons were statistically significant." (page 290 of 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:

The following tables (Table 10 for male rats, Table 11 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived to the end of the interval.

Table 10. Summary of Male Rat Survival (daily dose)

Period (Weeks)	Vehicle Control	Low 30 mg/kg	Medium 100 mg/kg	High 200 mg/kg
0-52	2/55 96.4%	3/55 94.5%	5/55 90.9%	4/65 93.8%
53-78	5/53 87.3%	3/52 89.1%	3/50 85.4%	5/61 86.1%
79-91	2/48 83.6%	10/49 70.9%	8/47 70.9%	17/56 60.0%
92-104	7/46 70.9%	6/39 60.0%	11/39 50.9%	9/39 46.2%
Terminal 105-106	39	33	28	30

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

Table 11. Summary of Female Rat Survival (daily dose)

Period (Weeks)	Vehicle Control	Low 30 mg/kg	Medium 100 mg/kg	High 200 mg/kg
0-52	0/55 100%	3/55 94.5%	4/55 92.7%	11/65 83.1%
53-78	5/55 81.9%	5/52 85.4%	11/51 72.7%	12/54 64.6%
79-91	4/50 83.6%	4/47 78.2%	2/40 69.1%	12/42 46.1%
92-105	8/46 69.1%	12/43 56.4%	10/38 50.9%	6/30 36.9%
Terminal 105-106	38	31	28	24

[1] number deaths / number at risk
[2] per cent survival to end of period.
[3] Includes one 50 mg/kg-dosed female which was sacrificed in poor condition on Day 4 and was replaced by an animal from the spare group.

The statistical significances of the tests of differences in rat survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table 12 below. As noted in Section 1.3.1.1 above, the Wilcoxon test weights events more equally over time and thus will be more sensitive to earlier separation of mortality than will be the log rank test.

Table 12. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Homogeneity over all 4 Groups	0.0400	0.0477	0.0004	<0.0001
Trend over Groups Vehicle to High	0.0070	0.0096	<0.0001	<0.0001
Comparison of Vehicle and High	0.0060	0.0072	<0.0001	<0.0001

Kaplan-Meier estimated survival curves across dose groups for each gender in each study are given in Figures A.1.1-A.1.2 of Appendix 1. From Table 12 above there is evidence of differences in survival curves in both rat genders, but the result is particularly strong in female rats (Male rat Logrank p=0.0400, Wilcoxon p=0.0477, Female Logrank p=0.00040, Wilcoxon p < 0.0001). From the Kaplan-Meier product limit survival curves in Figures A.1.1 and A.1.2 one can see that in male rats, and especially female rats, there is a evidence of a difference between general increase in mortality with increasing dose consistent with the above and even more with the more powerful test of trend in dose (Male rat trend: Logrank p=0.0070, Wilcoxon p=0.0072, Female rats trend both p<0.0001). Results for the pairwise comparison between the vehicle and high dose group were similar to the results for trend (Male rats: Logrank p=0.0060, Wilcoxon p=0.0072, Female rats: both Logrank and Wilcoxon p < 0.0001).

Appendix 2 includes an experimental Bayesian analysis that allows nonproprotional hazards, by defining a period wise constant hazard function for each dose. It also suggests an increasing hazard over increasing dose, though sometimes differences were sometimes relatively small.

Tumorigenicity analysis:

Table 13 below lists all tumors that have a potentially statistically significant trend or pairwise comparison with the vehicle controls (i.e., any $p \leq 0.05$). Appendix 3 includes complete incidence tables, including results with mice, and the corresponding poly-k tests of trend and pairwise differences. Applying the Haseman-Lin-Rahman rules, the test for trend in rare tumors (incidence ≤ 1%) could be considered statistically significant if the observed p-value is 0.025 or less, while common tumors (incidence > 1%) would be considered as statistically significant if the observed p-value is 0.005 or less. The pairwise test between the high dose group and the vehicle could be considered statistically significant if the observed p-value is 0.05 or less in rare tumors, and 0.01 or less in common tumors. Including the pairwise comparisons between the vehicle controls and the low and medium dose groups would be expected to increase the type I error rate above the nominal rough 10%. In this analysis, we use the incidence in the vehicle group to define whether the tumor is classified as rare or common.

Table 13. Potentially Statistically Significant Neoplasms

Organ	Incidence				p-values			
Tumor	Veh	Low	Med	High	Trend	Veh vs Hi	Veh vs Med	Veh vs Low
Female Rats								
OVARIES								
B-GRANULOSA CELL TUMOUR	0	0	2	4	0.0038	0.0326	0.1707	.
Sertoli/Granulosa Cell Tumor	1	0	2	4	0.0142	0.1083	0.3747	0.4891

Complete incidence tables are given in Tables A.3.2 and A.3.3 of Appendix 3. Using the incidence in the vehicle control group to define the rarity of the neoplasms need to apply the Haseman-Lin-Rahman adjustement for the multiplicity of tests, only the tests of trend in benign granulosa cell tumors in female rats were statistically significant (i.e., p = 0.0038 < 0.025), and the pairwise comparison of the high dose group to control (i.e., p = 0.0326 < 0.05). The test of trend in pooled Sertoli and granulose cell tumors

was nominally statistically significant (i.e., p = 0.0142), but if we strictly follow the Haseman-Lin-Rahman rules, this p > 0.005, and hence would be labeled as not statistically significant. No other tests in rats even achieved the usual 0.05 level of significance.

3.2.2. Study BOI287/042668 Carcinogenicity study by oral gavage administration to CD-1 mice

STUDY DURATION: 104/105 Weeks.
DOSING START DATE: June 19, 2003
TERMINAL SACRIFICE (NECROPSY) DATES: June 20-28, 2005.
STUDY ENDING DATE (Last Dosing): June 27, 2005.
MOUSE STRAIN: Crl:CD-1 [ICR]BR VAF.
ROUTE: Gavage

The Sponsor's report, apparently citing the testing facility, states that: "The dosages used in this study (0, 30, 100 and 200 mg/kg/day) were selected in conjunction with the Sponsor with reference to previous work with this compound performed in these laboratories (b) (4) Report Numbers: BOI 267/020215 and BOI 276/032942). In the study BOI 276/020215, a 13-week study, 300 mg/kg/day was associated with mortality due to haemorrhage, an expected consequence of the pharmacology of BIBR 1048 MS, precluding the administration of this dosage for a long term study. The low-dose of 30 mg/kg/day was intended to provide information on the No Observed Toxic Effect Level. It was estimated that this dose would result in plasma concentrations of BIBR 1048 MS in excess (>2-fold) of those expected in maximum in man at the clinical trial doses employed to date (up to 300 mg per patient). A larger group size was used for the highest dosage group on the current study to allow for the potentially higher mortality due to the pharmacological effects of BIBR 1048 MS.

The Sponsor's report further indicates that the testing facility (b) (4) has a policy of improvement of animal welfare which includes permitting social interaction through multiple housing of mice. As there was no indication on a previous study ((b) (4) Report no. BOI276/032942) that multiple housing would cause any changes in behaviour or survival, the animals were initially housed as follows: For Weeks 1 to 71 the animals were housed three of one sex per cage (except for the last cage of Group 1 Satellites which were housed two of one sex per cage) unless this number was reduced by mortality or isolation. Due to problems associated with fighting, from Week 72 onwards male animals were housed individually." (page 31 of report)

3.2.1.1. Sponsor's Results and Conclusions

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

Survival analysis:

According to the Sponsor, the number of animal deaths during the study, up to and including Week 104 for males and Week 102 for females, were analysed by logrank tests for a trend across the groups. Results were summarized as follows:

"Males
The trend test was not statistically significant when all groups were included in the analysis (p=0.993). None of the pairwise comparisons were statistically significant.

"Females
The trend test was not statistically significant when all groups were included in the analysis (p=0.097). None of the pairwise comparisons were statistically significant." (page 280 of report)

Tumorigenicity analysis:

The Sponsor reported the results of a Peto analysis using the Haseman-Lin-Rahman rules to adjust for multiplicity in Table 14 below, transcribed from page 300 of the Sponsor's report.

Table 14. Interpretation of *p*-values using Common/Rare classification

Sex	Tissue	Tumour	Historical Incidence (1)	Class	Statistical comparison and pvalue		Threshold (2)	Significant
Males	Skin	Benign fibroma, malignant fibrosarcoma and malignant sarcoma combined	9/452 (2%)	Common	Trend test Including all groups	0.045	0.005	No
					Pairwise control vs Group 4	0.035	0.01	No
Males	Haemato-poietic	Malignant pleomorphic lymphoma	18/452 (4%)	Common	Pairwise Control *vs* Group 3	0.032	0.01	No
Females	Lungs/ Bronchi	Benign bronchioloalveolar adenoma	101/449 (22%)	Common	Pairwise Control *vs* Group 3	0.045	0.01	No

(1) Based on control animals from six similar studies.
(2) Based on FDA Draft Guidance (FDA 2001).

These were reported verbally by the Sponsor as follows:

"Males
Skin
For benign fibroma, malignant fibrosarcoma and malignant sarcoma combined, the trend test was statistically significant when all groups were included in the analysis (p=0.045). Upon exclusion of the 200 mg/kg/day dosage group, the trend test was no longer statistically significant (p=0.176). The pairwise comparison between the control group

and the 200 mg/kg/day dosage group was statistically significant (p=0.035). None of the other pairwise comparisons were statistically significant.

Haematopoietic tumour
For malignant pleomorphic lymphoma, the trend test was not statistically significant when all groups were included in the analysis (p=0.360). The pairwise comparison between the control group and the 100 mg/kg/day dosage group was statistically significant (p=0.032). None of the other pairwise comparisons were statistically significant.

"Females
Lungs/Bronchi
For benign bronchioloalveolar adenoma, the trend test was not statistically significant when all groups were included in the analysis (p=0.349). The pairwise comparison between the control group and the 100 mg/kg/day dosage group was statistically significant (p=0.045).

"None of the other pairwise comparisons were statistically significant." (page 34 of report).

The Sponsor reports that, using the Haseman-Lin-Rahman rules no tumors sites showed significant differences.

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:

The following tables (Table 15 for male mice, Table 16 for female mice) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived to the end of the interval.

Table 15. Summary of Male Mice Survival (Eslicarbazepine: daily dose)

Period (Weeks)	Vehicle Control	Low 30 mg/kg	Medium 100 mg/kg	High 200 mg/kg
0-52	13/54 75.9%	5/54 90.7%	9/54 83.3%	8/63 87.3%
53-78	8/41 61.1%	11/49 70.4%	7/45 70.4%	12/55 68.2%
79-91	6/33 50.0%	8/38 55.6%	10/38 51.8%	11/43 50.8%
92-103	7/27 37.0%	9/30 38.9%	8/28 37.0%	10/32 34.9%
Terminal 103-107	20	21	20	22

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

Table 16. Summary of Female Mice Survival (daily dose)

Period (Weeks)	Vehicle Control	Low 30 mg/kg	Medium 100 mg/kg	High 200 mg/kg
0-52	2/54 96.3%	4/54 92.6%	5/54 90.7%	6/63 90.5%
53-78	12/52 74.1%	9/50 75.9%	12/49 68.5%	17/57 63.5%
79-91	6/40 63.0%	10/41 57.4%	9/37 51.8%	8/40 50.8%
92-104	12/34 40.7%	9/31 40.7%	13/28 27.8%	13/32 30.2%
Terminal 104-105	22	22	15	19

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

The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table 17 below. As noted in Section 1.3.1.2 above the logrank test will be more sensitive to later separation of mortality than will be the Wilcoxon test. Note that unlike the results in rats there is no strong evidence of a trend or other dose related differences in survival.

Table 17. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Homogeneity over all 4 Groups	0.9519	0.8547	0.3065	0.3258
Trend over Groups Vehicle to High	0.8735	0.9560	0.0950	0.0814
Comparison of Vehicle and High	0.9227	0.6554	0.1563	0.1202

Figures A.1.3 and A.1.4 in Appendix 1, seem to suggest a roughly increasing mortality with decreasing dose in male mice and a vague increasing mortality with

increasing dose in female mice. In male mice the vehicle dose group has generally the highest or close to highest mortality, with the medium dose group next, but closely intertwined with the high dose group. Finally the low dose group seems to roughly have the lowest mortality. In female mice the survival curves for the high dose group and the medium dose group are rather intertwined, with lower mortality than the similarly intertwined curved for the low dose group and the vehicle group. However, no tests of lack of overall homogeneity, pairwise homogeneity, or trend were statistically significant at the usual 0.05 level in either gender, (Males: all $p \geq 0.6554$, Females: all $p \geq 0.0814$).

Appendix 2 includes an experimental Bayesian analysis that allows nonproprotional hazards, by defining a period wise constant hazard function for each dose. It suggests there are no strong dose related differences in hazard.

Tumorigenicity analysis:

Table 18 below lists tumors that have any p-value less than 0.05. Using the incidence in the vehicle control group to define the rarity of the neoplasms, both of the tumor in the table would be classified as rare tumors in mice. In male mice the pairwise comparison of the medium dose group to the control in pleomorphormic h'poietic tumors was statistically significant (i.e., $p = 0.0217 < 0.05$), however, there was no other evidence of a dose related trend (i.e., trend $p = 0.3035$ and both other pairwise comparisons have $p \geq 0.1249$). In male mice the pairwise comparison of the high dose group to vehicle in pooled fibroma, fibrosarcoma, and sarcoma of the skin was statistically significant (i.e., $p = 0.0182 < 0.05$). However, for this tumor, although close to significance, using the Haseman-Lin-Rahman rules, no other test was statistically significant (i.e., trend $p = 0.0418 > 0.025$ and both other pairwise comparisons have $p = 0.0555, 0.0948 > 0.05$). Note that no other pairwise tests or tests of trend in achieved statistical significance using the nominal 0.05 level.

Table 18. Potentially Statistically Significant Neoplasms

Organ	Incidence				p-values			
Tumor	Veh	Low	Med	High	Trend	Veh vs Hi	Veh vs Med	Veh vs Low
Male Mice								
H'POIETIC TUMOUR								
PLEOMORPHIC LYMPHOMA	0	3	5	2	0.3035	0.2636	0.0217	0.1249
SKIN (PROTOCOL)								
Fibroma/Fibrosarcoma/Sarcoma	0	4	3	6	0.0418	0.0182	0.0948	0.0555

Complete incidence tables are given in Tables A.3.4 and A.3.5 of Appendix 3.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see section 1.1 above.

APPENDICES:

Appendix 1. Survival Analysis

Simple summary life tables in mortality are presented in the report (Tables 10, 11, 15, and 16), above. Kaplan-Meier estimated survival curves across dose groups for each gender in each species are displayed in Figures A.1.1-A.1.4 below. These plots include 95% confidence intervals around each curve (colored area around each curve). The plots are also supported by tests of homogeneity over the four dose groups with vehicle controls, simple tests of trend in survival, and finally a pairwise comparison between the high dose and the vehicle groups. The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table A.1.1, below. One might note that the log rank tests tends to put higher weight of later events and thus is more sensitive to later differences in survival. The Wilcoxon test tends to weight events more equally and thus is more sensitive to earlier differences in survival. If the survival curves track each other, then the proportional hazards assumption will be tenable, and the log rank test will be the most powerful.

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

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Rats				
Homogeneity over all 4 Groups	0.0400	0.0477	0.0004	<0.0001
Trend over Groups Vehicle to High	0.0070	0.0096	<0.0001	<0.0001
Comparison of Vehicle and High	0.0060	0.0072	<0.0001	<0.0001
Mice				
Homogeneity over all 4 Groups	0.9519	0.8547	0.3065	0.3258
Trend over Groups Vehicle to High	0.8735	0.9560	0.0950	0.0814
Comparison of Vehicle and High	0.9227	0.6554	0.1563	0.1202

From Table A.1.1 above there is evidence of differences in survival curves in rats, particularly female rats (Male rats Logrank p=0.0400, Wilcoxon p=0.0477, Female rats Logrank p=0.00040, Wilcoxon p < 0.0001). From the Kaplan-Meier product limit survival curves in Figures A.1.1 and A.1.2 below one can see that in male rats, and especially female rats, there is a evidence of a difference between general increase in mortality with increasing dose consistent with the above and even more with the more powerful test of trend in dose (Male rats trend: Logrank p=0.0070, Wilcoxon p=0.0072, Female rats trend both p<0.0001). Results for the pairwise comparison between the vehicle and high dose group were similar to the results for trend. In mice there was no strong evidence of heterogeneity in survival over dose, although from Figures A.1.3 and A.1.4 below, in female mice there seems to be suggestion of a roughly increasing mortality with decreasing dose. Note that in male mice the vehicle dose group has generally the highest or close to highest mortality, with the medium dose group next, but closely intertwined with the high dose group. Finally the low dose group seems to roughly have the lowest mortality. In female mice the survival curves for the high dose group and the medium dose group are rather intertwined, with lower mortality than the

similarly intertwined curved for the low dose group and the vehicle group. However, no tests of lack of overall homogeneity, pairwise homogeneity, or trend were statistically significant at the usual 0.05 level in either gender, (Males: all $p \geq 0.6554$, Females: all $p \geq 0.0814$).

Figures A.1.1 and A.1.2, below, display these Kaplan-Meier estimated survival curves for the two genders in Rats.

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

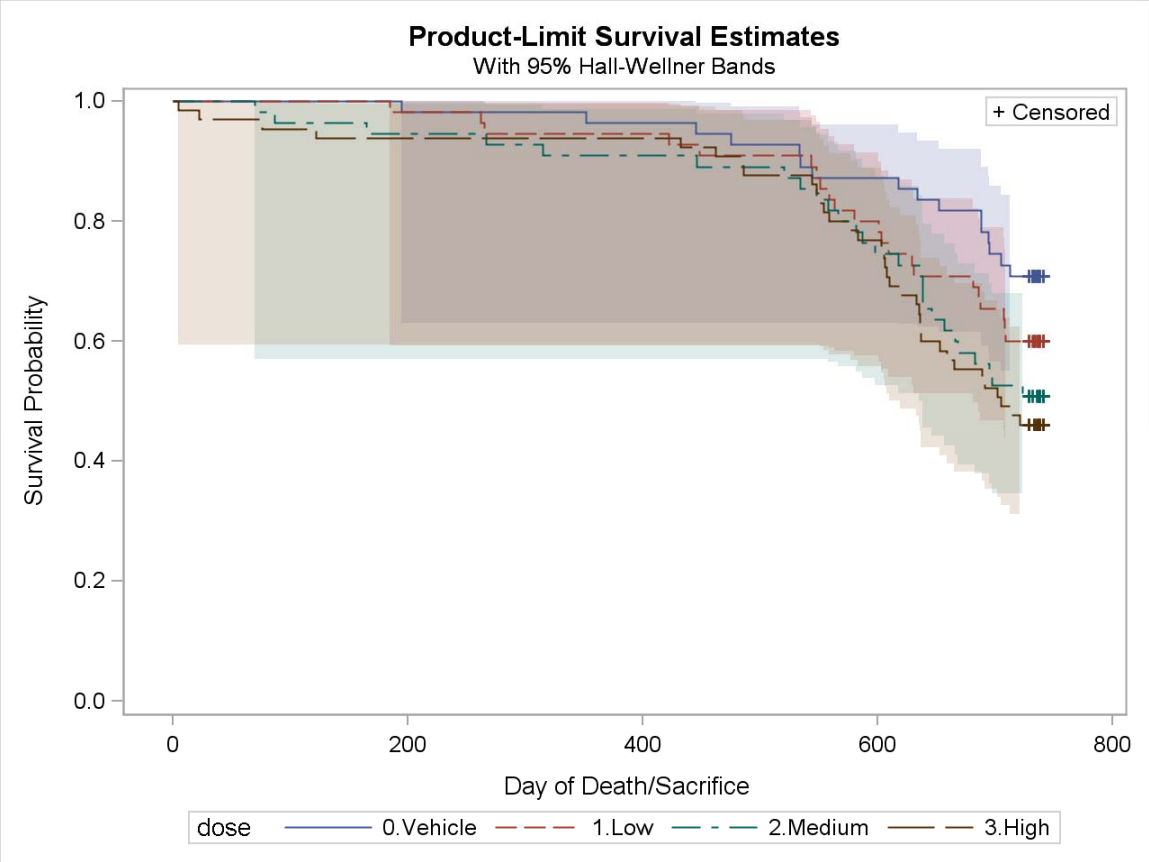
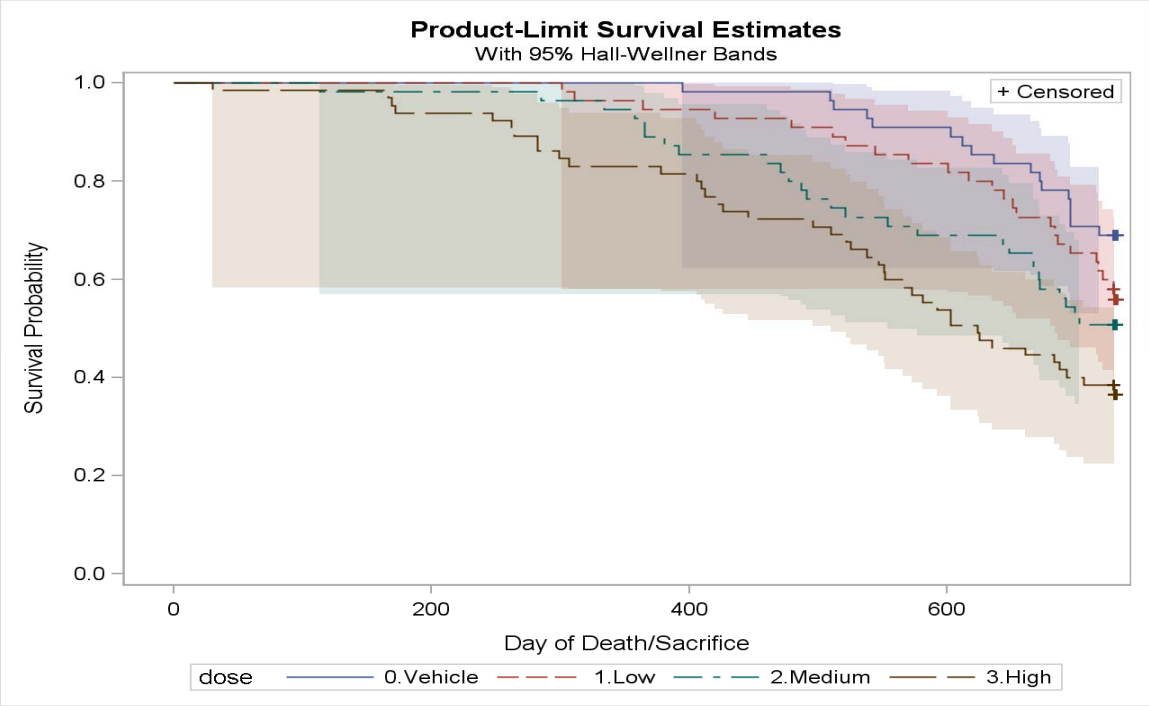


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



Figures A.1.3 and A.1.4, below, display these Kaplan-Meier estimated survival curves for the two genders in Mice

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

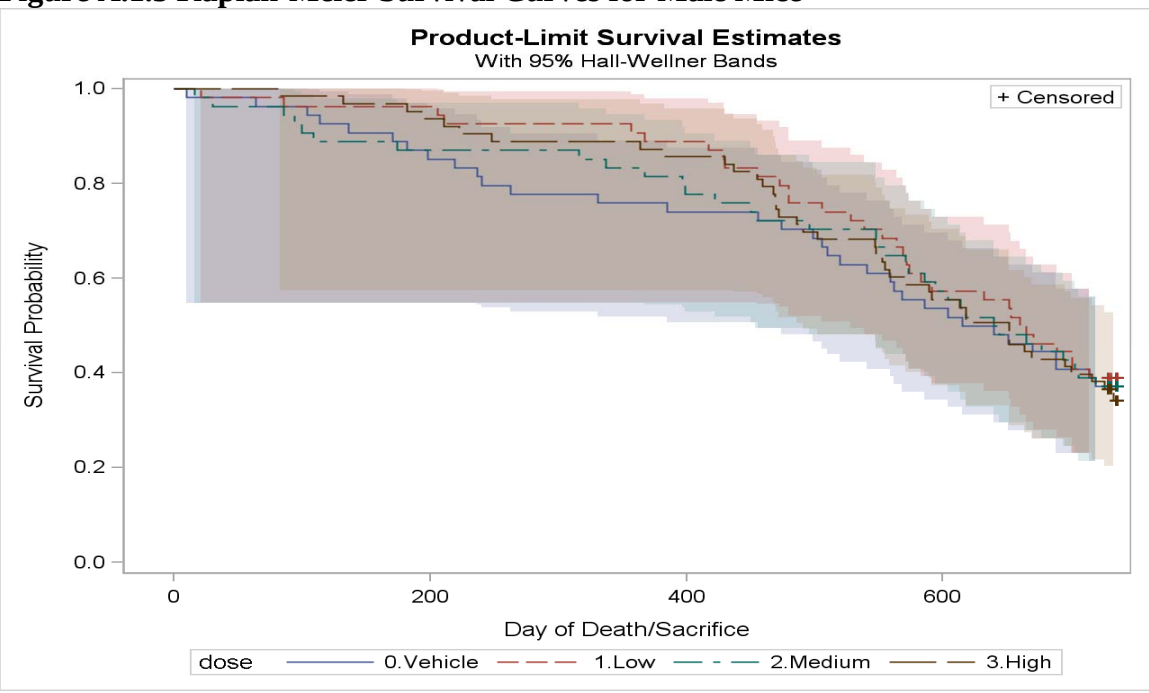
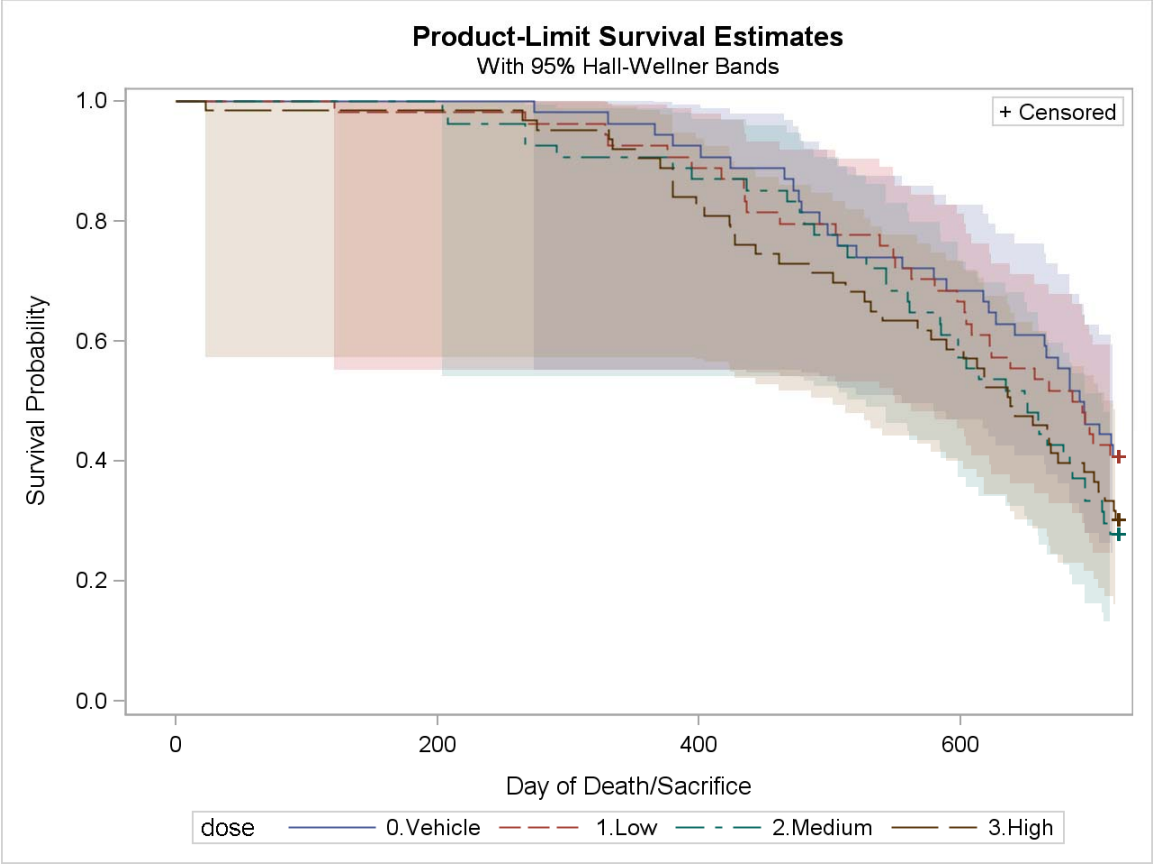


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



Appendix 2. Bayesian Analysis of Survival

Let *S(t)* be the survival function, i.e., with T denoting the random survival time,

$S(t) = Pr(T > t)$,

and *f(t)* the density of T. The instantaneous hazard function is $h(t) = f(t)/S(t)$ with cumulative hazard up to time t:

$$H(t) = \int_0^t h(u)\,du$$

So $f(t) = h(t)\,S(t)$. Also $log(S(t)) = -H(t)$, so $S(t) = e^{-H(t)}$. Then $f(t) = h(t)\,e^{-H(t)}$.

The standard Cox regression form of the proportional hazards model for survival specifies the hazard function:

$h(t \mid x) = h_0(t)\,exp(x^t\beta)$,

where $h_0(t)$ is the same across treatment groups. Then dose group only enters $h(t \mid x)$ through $exp(x^t\beta)$, and for each treatment group d, $H(t)$ is proportional to $exp(x^t\beta_i)$. Since $log(S(t)) = -H(t)$, the corresponding survival curves will tend to track each other without intersecting.

It is clear from the plots in Section 1 that several of the observed survival curves do intersect. A typical frequentist approach in such a circumstance is to introduce time dependent covariates to adjust for such intersections. Arguably a more sensible approach is to allow the estimated baseline hazard $h_0(t)$ to differ across treatment groups, i.e., a baseline hazard $h_{0d}(t)$ for data from the dth dose group. Perhaps the simplest Bayesian model would postulate a within interval constant baseline hazard, but a different hazard for each treatment group. That is, suppose the time axis can be partitioned as $(a_1=0, a_2]$, $(a_2, a_3]$, . . . , $(a_T,a_{T+1}]$. Assume a constant baseline hazard λ_{dj} in the jth time interval $(a_j,a_{j+1}]$, with observations from treatment group d, out of a total of g groups.

Let t_i = time to failure or censoring for the ith subject and suppose it is in the interval $(a_{j-1},a_j]$. So the integrated cumulative baseline hazard for this subject can be written as:

$$H_{od}(t_i) = \int_0^{t_i} h_{0d}(u)\,du = \{\sum_{k=1}^{j-1} \lambda_{dk}(a_k - a_{k-1}) + \lambda_{dj}(t_i - a_{j-1})\},$$

with instantaneous hazard $h_{od}(t_i) = \lambda_{dj}$. Note that the cumulative hazard will be represented as piece-wise linear function, with the cumulative hazard increasing at the constant rate λ_{dj} within each interval $(a_{j-1},a_j]$

The likelihood for subject i in dose group d can then be written as:

$$L_i(\lambda,\beta) \propto \begin{cases} e^{-H_{0d}(t_i)} & \text{if ith subject is censored at time } t_i \\ \lambda_{dj} e^{-H_{0d}(t_i)} & \text{if ith subject fails at time } t_i \end{cases}$$

Because this looks like a sample of exponential interarrival times we would expect the simple fail/not fail distributions to correspond to Poisson random variables.

For subject i censored or failed at time t_i, let $\gamma_{idk} = \begin{cases} \lambda_{dk}(a_k - a_{k-1}) & for\ t_i > a_k \\ \lambda_{dj}(t_i - a_{j-1}) & for\ a_{j-1} \le t_i < a_j \\ 0 & otherwise \end{cases}$

For intervals above a_j, the term $\gamma_{idk} = 0$, so for those intervals $\exp(-\gamma_{idk})$ does not contribute to the product. Then $S_d(t) = e^{-H(t)} = \prod_{k=1}^{T} \exp(-\gamma_{idk})$. Thus, for subject i in group d, the likelihood can also be written as:

$$L_i(\lambda, \beta) \propto \begin{cases} \prod_{k=1}^{T} \exp(-\gamma_{idk}) & if\ ith\ subject\ is\ censored\ at\ time\ t_i \\ \gamma_{idj} \prod_{k=1}^{T} \exp(-\gamma_{idk}) & if\ ith\ subject\ fails\ at\ time\ t_i \end{cases}$$

Note this corresponds to the likelihood of T independent Poisson random variables with mean γ_{dij} where all responses are zero except at time j with the occurrence of a failure in the jth interval $(a_{j-1}, a_j]$. Since all responses are 0 or 1, this is only a computational convenience but allows estimation of the appropriate parameters. As an aside, note that it would be easy to incorporate other individual level covariates besides dose in a the same manner as a typical semi-parametric cox regression, i.e. for the ith subject with covariates x_i, merely by replace each λ_{dj} by $\lambda_{dj}\ exp(\beta^t x_i)$. However that is not done in this analysis.

The time intervals used for the baseline hazards in this analysis match those used in Tables 10, 11, 15, and 16. These are long intervals, but for robustness of results we need to have a sufficient number of observations to estimate the within dose, within time period, probability of mortality. Sacrifice or accidental death is treated as a reduction in the risk set, but not as a mortality event. A gamma prior on the within treatment group, within time period, hazard would be skewed to the right and would seem to be an appropriate choice of family for the baseline hazards λ_{dj}.

To reflect the expectation of an increasing hazard, for period j, j=1,…,4, we specify a gamma prior on the λ_{dj} with location 0.25 and scale parameter 25*j. This implies a prior scaled hazard with mean 100*j and variance 400*j. This would seem to be a relatively noninformative prior. Tables A.2.1-A.2.4, below, summarize the estimated posterior distributions of the treatment parameters. Here, time in study was measured in weeks. Since the time intervals are of different lengths, and to increase readability in the tables the mean in the tables below correspond to $100*\lambda_{dj}$, i.e., 100 times the hazard. This is the Poisson mean times 100 divided by the length of the interval, and might be called "normalized hazard means". The standard deviation of the values and the 0.25%, median (i.e. 50%), and 97.5% quantiles are also presented. One measure of difference between doses is, within each time period, to compare this mean to the average of the other three dose groups. For assessing overall treatment differences

this is the most interesting measure. The mean difference, its standard deviation, and the corresponding quantiles are also presented. Note that the actual probability that the differences are in the range between the 2.5% and 97.5% percentiles is 0.95, defining a so-called 95% credible interval. Thus, if 0 is not in that interval, we know that the probability that each parameter is "close" to the mean of the remaining parameters is less than 0.05, providing relatively strong evidence that they are different.

Table A.2.1 Posterior Summaries of Treatment Parameters in Male Rats

		Normalized mean		Quantiles			Difference		Quantiles		
Weeks	Dose	Mean	std dev	2.5%	50%	97.5%	mean	sd diff	2.5%	50%	97.5%
0-52	0	0.0112	0.0075	0.0016	0.0096	0.0301	-0.0095	0.0094	-0.0265	-0.0102	0.0114
	30	0.0163	0.0090	0.0037	0.0147	0.0378	-0.0027	0.0105	-0.0210	-0.0037	0.0204
	100	0.0271	0.0118	0.0092	0.0254	0.0547	0.0117	0.0128	-0.0096	0.0105	0.0407
	200	0.0187	0.0090	0.0055	0.0173	0.0401	4.91E-4	0.0106	-0.0181	-4.372E-4	0.0240
53-78	0	0.0556	0.0241	0.0188	0.0522	0.1117	0.0051	0.0275	-0.0424	0.0027	0.0661
	30	0.046	0.0223	0.0134	0.0425	0.0995	-0.0077	0.0262	-0.0534	-0.0100	0.0503
	100	0.0475	0.0230	0.0135	0.0438	0.1011	-0.0058	0.0266	-0.0514	-0.0083	0.0536
	200	0.0581	0.0232	0.0221	0.0548	0.1116	0.0084	0.0268	-0.0386	0.0063	0.0666
79-91	0	0.0477	0.0316	0.0070	0.0409	0.1258	-0.2283	0.0551	-0.3361	-0.2285	-0.1180
	30	0.2163	0.0714	0.1006	0.2079	0.3761	-0.0035	0.0819	-0.1473	-0.0096	0.1715
	100	0.2705	0.0808	0.1369	0.262	0.4519	0.0687	0.0897	-0.0888	0.0621	0.2633
	200	0.3412	0.0846	0.1963	0.3342	0.5248	0.163	0.0926	-0.0034	0.1565	0.3604
92-term-	0	0.1916	0.0715	0.0783	0.1824	0.3572	-0.0563	0.0884	-0.2189	-0.0602	0.1314
inal	30	0.1902	0.0755	0.0733	0.1802	0.3641	-0.0581	0.0912	-0.2217	-0.0636	0.1358
	100	0.256	0.0946	0.1046	0.2447	0.4734	0.0296	0.1059	-0.1558	0.0210	0.2613
	200	0.2974	0.0978	0.1386	0.2859	0.5187	0.0848	0.1081	-0.1036	0.0769	0.3191

Table A.2.2 Posterior Summaries of Treatment Parameters in Female Rats

		Normalized mean		Quantiles			Difference		Quantiles		
Weeks	Dose	Mean	std dev	2.5%	50%	97.5%	mean	sd diff	2.5%	50%	97.5%
0-52	0	0.0013	0.0025	1.397E-9	2.127E-4	0.0086	-0.0314	0.0076	-0.0474	-0.0311	-0.0175
	30	0.0163	0.0091	0.0036	0.0146	0.0383	-0.0114	0.0113	-0.0312	-0.0122	0.0131
	100	0.0314	0.0126	0.0120	0.0298	0.0604	0.0089	0.0139	-0.0147	0.0076	0.0398
	200	0.0502	0.0150	0.0253	0.0487	0.0837	0.0338	0.0159	0.0060	0.0326	0.0684
53-78	0	0.0535	0.0233	0.0181	0.0500	0.1085	-0.0535	0.0310	-0.1111	-0.0548	0.0123
	30	0.0569	0.0248	0.0193	0.0533	0.1144	-0.0490	0.0321	-0.1082	-0.0506	0.0190
	100	0.1127	0.0373	0.0518	0.1086	0.1968	0.0254	0.0414	-0.0473	0.0223	0.1147
	200	0.1515	0.0417	0.0814	0.1476	0.2441	0.0771	0.0450	-0.0021	0.0740	0.1746
79-91	0	0.0873	0.0423	0.0252	0.0804	0.1875	-0.0957	0.0583	-0.2058	-0.0973	0.0260
	30	0.1158	0.0508	0.0390	0.1082	0.235	-0.0577	0.0643	-0.1765	-0.0608	0.0793
	100	0.1085	0.0522	0.0311	0.1005	0.2328	-0.0674	0.0653	-0.1859	-0.0715	0.0727
	200	0.3246	0.0967	0.1632	0.3159	0.5378	0.2207	0.1007	0.0483	0.2125	0.4409
92-term-	0	0.2326	0.0808	0.1031	0.2241	0.4148	-0.0771	0.1014	-0.2641	-0.0812	0.1315
inal	30	0.3518	0.1053	0.1753	0.3418	0.5869	0.0820	0.1199	-0.1359	0.0757	0.3371
	100	0.3089	0.1076	0.1352	0.2965	0.5511	0.0246	0.1212	-0.189	0.0153	0.2835
	200	0.2684	0.1061	0.1014	0.2549	0.5089	-0.0294	0.1204	-0.2389	-0.0397	0.2344

In male rats the only 95% credible interval for differences in hazard that does not contain zero is the interval for the difference between the vehicle and the other groups during the 79-91 week period. The interval is negative, indicating that the vehicle has a lower hazard than the average of the other three treatments. In female rats the interval for the difference in hazard between the vehicle and the other groups during the 0-52 week period is completely negative, indicating that the vehicle has significantly lower hazard during this period. However, the credible interval for the differences in hazard between the high dose group and the average of the other groups is positive in both the 79-91 and 92-terminal periods, indicating higher hazard in the high dose group during those periods. From a Bayesian perspective, the frequentist analysis in Appendix 1 tends to detect arguable differences. However, while these results are consistent with the

frequentist analysis, one may question whether comparing the hazard of a dose group to the mean of the remaining groups is an appropriate measure. This is an issue for further research.

Table A.2.3 Posterior Summaries of Treatment Parameters in Male Mice

		Normalized mean		Quantiles			Difference		Quantiles		
Weeks	Dose	Mean	std dev	2.5%	50%	97.5%	mean	sd diff	2.5%	50%	97.5%
0-52	0	0.0761	0.0209	0.0407	0.0741	0.1221	0.0333	0.0226	-0.0066	0.0318	0.0821
	30	0.0331	0.0132	0.0125	0.0313	0.0633	-0.0240	0.0167	-0.0547	-0.0248	0.0113
	100	0.0574	0.0178	0.028	0.0555	0.0974	0.0084	0.0201	-0.0271	0.0071	0.0515
	200	0.0378	0.0132	0.0169	0.0363	0.0679	-0.0177	0.0166	-0.0485	-0.0185	0.0178
53-78	0	0.1176	0.0408	0.0523	0.1125	0.2103	-0.0154	0.0469	-0.0994	-0.0187	0.0854
	30	0.1281	0.0399	0.0627	0.1234	0.218	-0.0014	0.0466	-0.0843	-0.0044	0.0985
	100	0.1118	0.0389	0.0494	0.1074	0.2	-0.0231	0.0456	-0.1048	-0.0259	0.074
	200	0.1591	0.0422	0.0876	0.1556	0.2521	0.0399	0.0482	-0.0475	0.0373	0.1414
79-91	0	0.2488	0.0918	0.1015	0.2381	0.4582	-0.0323	0.1067	-0.222	-0.0396	0.1972
	30	0.2528	0.0883	0.1104	0.2423	0.4519	-0.0270	0.1039	-0.2138	-0.0332	0.1943
	100	0.3289	0.1027	0.1589	0.3179	0.557	0.0745	0.1145	-0.1268	0.0658	0.3181
	200	0.2617	0.0856	0.1233	0.252	0.4552	-0.0151	0.1017	-0.1985	-0.0209	0.2025
92-term	0	0.306	0.1221	0.1144	0.2901	0.5847	-0.0804	0.1436	-0.3359	-0.0903	0.2339
	30	0.4257	0.139	0.2001	0.4106	0.7396	0.0792	0.1573	-0.2003	0.0696	0.4151
	100	0.2993	0.1189	0.1154	0.2827	0.5728	-0.0894	0.1415	-0.3432	-0.0999	0.2172
	200	0.4343	0.137	0.2106	0.4199	0.7379	0.0907	0.1553	-0.1877	0.0813	0.4255

Table A.2.4 Posterior Summaries of Treatment Parameters in Female Mice

		Normalized mean		Quantiles			Difference		Quantiles		
Weeks	Dose	Mean	std dev	2.5%	50%	97.5%	mean	sd diff	2.5%	50%	97.5%
0-52	0	0.0164	0.0091	0.0037	0.0147	0.0386	-0.0091	0.0112	-0.0289	-0.0100	0.0157
	30	0.0219	0.0106	0.0063	0.0202	0.0471	-0.0019	0.0122	-0.0232	-0.0029	0.0253
	100	0.0273	0.0120	0.0090	0.0256	0.0553	0.0053	0.0134	-0.0173	0.0040	0.0353
	200	0.0275	0.0111	0.0104	0.0261	0.0532	0.0056	0.0127	-0.0165	0.0047	0.0329
53-78	0	0.1342	0.0400	0.0673	0.1302	0.2236	-0.0270	0.0474	-0.1144	-0.0295	0.0737
	30	0.1343	0.0399	0.0680	0.1302	0.2227	-0.0268	0.0474	-0.1139	-0.0289	0.0715
	100	0.1503	0.0428	0.0787	0.1457	0.2454	-0.0055	0.0492	-0.0942	-0.0081	0.0994
	200	0.1988	0.0480	0.1153	0.1951	0.3011	0.0592	0.0536	-0.0376	0.0565	0.1719
79-91	0	0.1966	0.0729	0.0818	0.1871	0.365	-0.0827	0.0899	-0.2457	-0.0882	0.1097
	30	0.2643	0.0865	0.1243	0.2543	0.4565	0.0076	0.1001	-0.1703	0.0015	0.2221
	100	0.2884	0.0949	0.1336	0.2781	0.4982	0.0397	0.1063	-0.1487	0.0328	0.2688
	200	0.2851	0.0891	0.1376	0.2759	0.4826	0.0353	0.1017	-0.147	0.0296	0.2515
92-ter-	0	0.3809	0.1134	0.1932	0.3687	0.6352	-0.0625	0.1371	-0.3142	-0.0691	0.2266
	30	0.2907	0.0998	0.1288	0.2794	0.5153	-0.1828	0.1283	-0.423	-0.187	0.0829
	100	0.6026	0.1659	0.3228	0.5869	0.9666	0.2331	0.1783	-0.0804	0.2215	0.6172
	200	0.4369	0.1303	0.2211	0.4236	0.7274	0.0122	0.1502	-0.2575	0.0035	0.3292

Note that in no 95% credible interval for the differences between the hazard of each group and the average of the other groups exclude 0, suggesting there is no strong evidence of hazard differences among the treatment groups.

This is an experimental approach. The appropriateness of this analysis and its generalizations is a topic for further research.

These analyses were implemented in WinBUGS 1.4.3 (see Lunn et al, 2000)

Appendix 3. FDA Poly-k Analysis

Tables A.3.1-A.3.5 below, display the tumor incidence over the four dose groups, including a vehicle group and three actual dosing groups, as well as the p-values of the poly-k (here with k=3) tests of trend in dose and pairwise comparisons to the vehicle controls. The first p-value provides the results of the overall poly-k test of trend. The poly-k test modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The last three columns present the results of tests between the vehicle group and each of the high, medium, and low dose groups, respectively. Note that even the lack of strong evidence of differences in mortality in mice does not necessarily imply that tests of tumorigenicity do not need to be adjusted for differences in mortality.

As noted in the report, at the Society of Toxicological Pathology "town hall" meeting in June 2001 the poly-k modification of the Cochran-Armitage test of trend seemed to have been recommended over the so-called Peto tests. The tests used here are small sample exact tests, which assume all marginal totals are fixed, a debatable assumption. To adjust for the multiplicity of tests, tentatively, the Haseman-Lin-Rahman rules discussed in Section 1.3.1.4. of the report seem to apply.

Table A.3.1 below lists tumors that have any p-value less than 0.05 in either species. Applying the Haseman-Lin-Rahman rules to adjust for the multiplicity of statistical tests, the test for trend in rare tumors (incidence $\leq 1\%$) could be considered statistically significant if the observed p-value is 0.025 or less, while common tumors (incidence $> 1\%$) would be considered as statistically significant if the observed p-value is 0.005 or less. The pairwise test between the high dose group and the vehicle group could be considered statistically significant if the observed p-value is 0.05 or less in rare tumors, and 0.01 or less in common tumors. Including the pairwise comparisons between the vehicle controls and the low and medium dose groups would be expected to increase the type I error rate above the nominal rough 10%. In this analysis, we use the incidence in the vehicle control group to define whether the tumor is classified as rare or common.

In female rats, the test of trend in benign granulosa cell tumor was statistically significant, even following the Haseman-Lin-Rahman rules (i.e., $p=0.0038 < 0.025$), as was the pairwise test comparing the high dose group to control (i.e., $p=0.0326 < 0.05$). The test of trend in pooled Sertoli and granulose cell tumor was nominally statistically significant (i.e., $p = 0.0142$), but if we strictly follow the Haseman-Lin-Rahman rules, this $p > 0.005$, and hence would be labeled as not statistically significant. In male mice the pairwise comparison of the medium dose group to the control was also statistically significant (i.e., $p=0.0217 < 0.05$), however, there was no other evidence of a dose related trend (i.e., trend $p=0.3035$ and both other pairwise comparisons have $p \geq 0.1249$). In male mice the pairwise comparison of the high dose group to vehicle in pooled fibroma, fibrosarcoma, and sarcoma of the skin was also statistically significant (i.e., $p = 0.0182 < 0.05$). However, although close to significance, using the rules, no other test was statistically significant (i.e., trend $p = 0.0418 > 0.025$ and both other pairwise

comparisons have p = 0.0555, 0.0948 > 0.05). Note that no other pairwise tests or tests of trend in achieved statistical significance using the nominal 0.05 level.

Table A.3.1. Potentially Statistically Significant Neoplasms

Organ Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
Female Rats								
OVARIES								
B-GRANULOSA CELL TUMOUR	0	0	2	4	0.0038	0.0326	0.1707	.
Sertoli/Granulosa Cell Tumor	1	0	2	4	0.0142	0.1083	0.3747	0.4891
Male Mice								
H'POIETIC TUMOUR								
PLEOMORPHIC LYMPHOMA	0	3	5	2	0.3035	0.2636	0.0217	0.1249
SKIN (PROTOCOL)								
Fibroma/Fibrosarcoma/Sarcoma	0	4	3	6	0.0418	0.0182	0.0948	0.0555

Complete incidence tables are provided in Table A.3.2 to A.3.5 below:

Table A.3.2. Incidence of Neoplasms in Male Rats

Organ Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
ADRENAL								
B-CORTICAL ADENOMA	2	2	1	1	0.4790	0.5689	0.4396	0.6575
B-PHAEOCHROMOCYTOMA	2	2	1	2	0.5681	0.3796	0.4396	0.6575
Cortical Adenoma/Carcinoma	2	3	1	1	0.8499	0.5689	0.4396	0.4554
M-CORTICAL CARCINOMA	0	1	0	0	0.5211	.	.	0.4769
BRAIN								
B-GRANULAR CELL TUMOUR	0	0	1	1	0.2203	0.5455	0.4603	.
M-ASTROCYTOMA	1	0	1	0	0.6606	0.5455	0.7127	0.4769
M-MIXED GLIOMA	1	0	0	0	0.7483	0.5385	0.4531	0.4697
M-OLIGODENDROGLIOMA	0	1	0	1	0.3642	0.5513	.	0.4769
Oligodendroglioma/Mix. Glioma	1	1	0	1	0.5078	0.2931	0.4531	0.7226
EPIDIDYMIDES								
B-MESOTHELIOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
HAEMATOPOIETIC TUMOUR								
M-LYMPHOCITIC / LYMPHOBLASTIC LYMPHOMA	1	0	1	2	0.2343	0.5671	0.7049	0.4697
M-PLEOMORPHIC LYMPHOMA	1	0	1	0	0.6606	0.5455	0.7127	0.4769
Pooled Lymphomas	2	0	2	2	0.3851	0.3778	0.6327	0.7226
HARDERIAN GLANDS								
B-ADENOMA	0	0	0	2	0.0860	0.2943	.	.
JEJUNUM								
M-ADENOCARCINOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
M-LEIOMYOSARCOMA	0	0	0	1	0.2958	0.5455	.	.
KIDNEYS								
B-LIPOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
B-TUBULAR ADENOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
LIVER								
B-HEPATOCELLULAR ADENOMA	0	0	1	0	0.5211	.	0.4603	.
LUNGS/BRONCHI								
B-BRONCHIOLOALVEOLAR ADENOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
M-BRONCHIOLOALVEOLAR ADENOCARCINOMA	0	0	0	1	0.2958	0.5455	.	.
MAMMARY AREAS								
B-LIPOMA	0	0	0	1	0.2958	0.5455	.	.
B-MAMMARY FIBROADENOMA	0	1	0	0	0.5211	.	.	0.4769
MESENTERIC LYMPH NODE								
B-HAEMANGIOMA	5	8	3	6	0.7399	0.3914	0.5515	0.2100

Table A.3.2. (cont.) Incidence of Neoplasms in Male Rats

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
PANCREAS								
Acinar/Islet/Mixed Cell Adenoma	2	2	0	4	0.2685	0.4286	0.7127	0.6575
B-ACINAR CELL ADENOMA	0	1	0	0	0.5211	.	.	0.4769
B-ISLET CELL ADENOMA	2	1	0	3	0.3444	0.5870	0.7127	0.4649
B-MIXED CELL ADENOMA	0	0	0	1	0.2958	0.5455	.	.
M-SCHWANNOMA	0	0	0	1	0.2958	0.5455	.	.
PARATHYROIDS								
B-ADENOMA	0	0	1	0	0.5211	.	0.4603	.
PAROTID S.G.								
M-SCHWANNOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
PENIS/PREPUCE								
M-SQUAMOUS CELL CARCINOMA	0	1	0	0	0.5211	.	.	0.4769
PITUITARY								
B-ADENOMA, PARS DISTALIS	19	17	17	19	0.7466	0.7036	0.5850	0.4973
B-ADENOMA, PARS INTERMEDIA	1	2	2	2	0.4449	0.5689	0.4396	0.4649
M-CARCINOMA, PARS DISTALIS	1	1	0	0	0.8252	0.5455	0.4603	0.7303
P.Dist.Adenoma/Carcinoma	20	18	17	19	0.8325	0.7824	0.4830	0.4919
PREPUTIAL GLANDS								
M-SQUAMOUS CELL CARCINOMA	1	1	1	1	0.4861	0.2943	0.7127	0.7303
PROSTATE								
B-ADENOMA	1	0	0	1	0.5551	0.2943	0.4603	0.4769
RECTUM								
B-FIBROMA	0	0	1	0	0.5211	.	0.4603	.
SKELETAL MUSCLE								
B-FIBROMA	0	0	1	0	0.5211	.	0.4603	.
SKIN								
B-BASAL CELL TUMOUR	0	0	2	1	0.2535	0.5455	0.2079	.
B-FIBROMA	1	2	1	2	0.4429	0.5569	0.7127	0.4649
B-KERATOACANTHOMA	3	2	1	1	0.8743	0.7582	0.6471	0.4554
B-SEBACEOUS CELL ADENOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769
B-SQUAMOUS CELL PAPILLOMA	1	0	1	1	0.4663	0.3007	0.7127	0.4769
Fibroma/Fibrosarcoma	1	3	1	3	0.3554	0.3667	0.7127	0.2839
M-FIBROSARCOMA	0	1	0	1	0.3578	0.5455	.	0.4769
M-SARCOMA NOS	0	1	0	0	0.5211	.	.	0.4769
M-SQUAMOUS CELL CARCINOMA	0	0	1	1	0.2203	0.5455	0.4603	.
Sq. Cell Papilloma/Carcinoma	1	0	2	2	0.2388	0.5777	0.4396	0.4769
TESTES								
B-INTERSTITIAL (LEYDIG)CELL ADENOMA	0	2	1	4	0.0573	0.0827	0.4603	0.2236
THYMUS								
B-THYMOMA (EPITHELIAL)	0	0	1	0	0.5211	.	0.4603	.
B-THYMOMA (LYMPHOID)	2	0	1	1	0.5918	0.5689	0.4396	0.7303
Epithelial Thymoma M&B	0	0	1	1	0.2203	0.5455	0.4603	.
M-MALIGNANT THYMOMA(EPITHELIAL)	0	0	0	1	0.2958	0.5455	.	.
THYROIDS								
B-C-CELL ADENOMA	1	4	3	4	0.2926	0.2413	0.2484	0.1501
B-FOLLICULAR CELL ADENOMA	1	5	4	4	0.3561	0.2337	0.1318	0.0789
C-Cell Adenoma/Carcinoma	2	4	4	5	0.2730	0.2980	0.2625	0.2926
Foll.Cell Adenoma/Carcinoma	2	5	4	4	0.4755	0.4182	0.2625	0.1767
M-C-CELL CARCINOMA	1	1	1	1	0.4861	0.2943	0.7127	0.7303
M-FOLLICULAR CELL CARCINOMA	1	0	0	0	0.7535	0.5455	0.4603	0.4769

Table A.3.3. Incidence of Neoplasms in Female Rats

Organ / Tumor	Incidence				p-values			
	Veh	Low	Med	High	Trend	Veh vs Hi	Veh vs Med	Veh vs Low
ADIPOSE TISSUE								
B-LIPOMA	0	0	0	1	0.2249	0.4353	.	.
ADRENAL								
B-CORTICAL ADENOMA	2	1	0	1	0.5977	0.4023	0.6627	0.4839
B-PHAEOCHROMOCYTOMA	1	0	1	1	0.3257	0.6840	0.6627	0.4894
Cortical Adenoma/Carcinoma	2	2	0	1	0.6840	0.4023	0.6627	0.6753
M-CORTICAL CARCINOMA	0	1	0	0	0.4320	.	.	0.4894
BRAIN								
B-GRANULAR CELL TUMOUR	0	1	1	0	0.4420	.	0.4235	0.4894
M-OLIGODENDROGLIOMA	0	1	0	0	0.4320	.	.	0.4894
CLITORAL GLANDS								
M-SQUAMOUS CELL CARCINOMA	0	0	1	0	0.4320	.	0.4167	.
HAEMATOPOIETIC TUMOUR								
M-PLEOMORPHIC LYMPHOMA	0	1	0	0	0.4320	.	.	0.4894
HEAD								
M-SQUAMOUS CELL CARCINOMA	0	0	0	1	0.2249	0.4353	.	.
HEART								
B-ENDOCARDIAL SCHWANNOMA	0	0	1	0	0.4320	.	0.4167	.
JEJUNUM								
B-LEIOMYOMA	0	1	0	0	0.4320	.	.	0.4894
M-LEIOMYOSARCOMA	1	0	0	0	0.7101	0.4353	0.4167	0.4894
KIDNEYS								
B-TUBULAR ADENOMA	0	0	1	0	0.4320	.	0.4167	.
LIVER								
B-HEPATOCELLULAR ADENOMA	0	0	1	1	0.1432	0.4353	0.4167	.
Hepato. Adenoma/Carcinoma	0	1	1	1	0.2333	0.4353	0.4167	0.4894
M-HEPATOCELLULAR CARCINOMA	0	1	0	0	0.4294	.	.	0.4894
MAMMARY AREAS								
Adenoma/Fibroad./Adenocarc.	20	13	13	10	0.8231	0.8845	0.5147	0.8727
B-MAMMARY ADENOMA	3	1	3	3	0.2185	0.5308	0.4913	0.6753
B-MAMMARY FIBROADENOMA	16	8	10	7	0.8129	0.8966	0.5369	0.9446
M-MAMMARY ADENOCARCINOMA	3	5	1	2	0.7337	0.3771	0.5560	0.3333
MESENTERIC LYMPH NODE								
B-HAEMANGIOMA	3	3	1	1	0.8047	0.5878	0.5560	0.6408
OVARIES								
B-GRANULOSA CELL TUMOUR	0	0	2	4	0.0038	0.0326	0.1707	.
B-SERTOLI CELL ADENOMA	1	0	0	0	0.7101	0.4337	0.4167	0.4891
Sertoli/Granulosa Cell Tumor	1	0	2	4	0.0142	0.1083	0.3747	0.4891
M-SCHWANNOMA	0	1	0	0	0.4320	.	.	0.4894
PANCREAS								
B-ISLET CELL ADENOMA	0	0	1	0	0.4320	.	0.4167	.
M-ACINAR CELL ADENOCARCINOMA	0	0	0	1	0.2249	0.4353	.	.
PAROTID S.G.								
B-ADENOMA	0	0	1	0	0.4320	.	0.4167	.
PITUITARY								
B-ADENOMA, PARS DISTALIS	35	33	25	27	0.5824	0.4404	0.4297	0.5791
M-CARCINOMA, PARS DISTALIS	1	1	0	1	0.5238	0.6840	0.4167	0.7419
P.Dist.Adenoma/Carcinoma	36	34	25	28	0.5796	0.4162	0.5089	0.5753
SKIN								
B-BASAL CELL TUMOUR	0	1	0	0	0.4320	.	.	0.4894
B-KERATOACANTHOMA	2	0	0	0	0.9172	0.6840	0.6627	0.7419
B-SQUAMOUS CELL PAPILLOMA	0	0	1	0	0.4337	.	0.4217	.
M-RHABDOMYOSARCOMA	0	0	1	0	0.4320	.	0.4167	.
TAIL								
M-HISTIOCYTIC SARCOMA	1	0	0	0	0.7101	0.4353	0.4167	0.4894
THORAX								
M-SCHWANNOMA	0	1	0	0	0.4320	.	.	0.4894
THYMUS								
B-THYMOMA (EPITHELIAL)	0	1	0	0	0.4320	.	.	0.4894
B-THYMOMA (LYMPHOID)	5	4	3	3	0.5968	0.4893	0.4436	0.4726
Epithelial Thymoma M&B	0	1	0	0	0.4320	.	.	0.4894

Table A.3.3. (cont.) Incidence of Neoplasms in Female Rats

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
THYROIDS								
B-C-CELL ADENOMA	4	3	1	0	0.9751	0.9039	0.6980	0.4760
B-FOLLICULAR CELL ADENOMA	0	0	1	1	0.1432	0.4353	0.4167	.
C-Cell Adenoma/Carcinoma	4	5	1	0	0.9875	0.9039	0.6980	0.4726
Foll.Cell Adenoma/Carcinoma	1	0	1	1	0.3257	0.6840	0.6627	0.4894
M-C-CELL CARCINOMA	0	2	0	0	0.6788	.	.	0.2368
M-FOLLICULAR CELL CARCINOMA	1	0	0	0	0.7101	0.4353	0.4167	0.4894
UTERINE CERVIX								
M-SCHWANNOMA	0	0	1	0	0.4353	.	0.4235	.
M-SQUAMOUS CELL CARCINOMA	0	1	0	0	0.4320	.	.	0.4894
M-STROMAL SARCOMA	1	0	0	0	0.7101	0.4302	0.4167	0.4842
UTERUS								
B-DECIDUOMA	0	0	0	1	0.2294	0.4419	.	.
B-ENDOMETRIAL ADENOMA	0	1	0	0	0.4320	.	.	0.4894
B-POLYP (STROMAL)	9	6	2	2	0.9791	0.9399	0.9171	0.6807
M-ADENOCARCINOMA	1	4	2	2	0.4120	0.4023	0.3821	0.1677
M-SCHWANNOMA	0	1	0	2	0.0869	0.1923	.	0.4894

Table A.3.4. Incidence of Neoplasms in Male Mice

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
ABDOMEN								
SARCOMA	1	0	0	0	0.7500	0.5088	0.4655	0.4815
ADRENALS								
CORTICAL ADENOMA	2	0	1	3	0.1909	0.5347	0.4474	0.7453
PHAEOCHROMOCYTOMA	0	1	2	0	0.4823	.	0.2123	0.4906
SUBCAPSULAR CELL ADENOMA	5	6	2	5	0.6855	0.4115	0.7268	0.4445
SUBCAPSULAR CELL CARCINOMA	0	1	0	0	0.5000	.	.	0.4906
BONE								
OSTEOMA	0	1	0	0	0.4960	.	.	0.4906
OSTEOSARCOMA	0	0	1	0	0.5000	.	0.4655	.
BRAIN								
SCHWANNOMA	0	1	0	0	0.4960	.	.	0.4906
GALL BLADDER								
PAPILLOMA	1	1	2	1	0.5357	0.2636	0.4474	0.7453
H'POIETIC TUMOUR								
LYMPHOCYTIC/LYMPHOBLASTIC LYMPHOMA	3	2	2	3	0.4839	0.6646	0.4208	0.4489
PLASMA CELL LYMPHOMA	0	1	0	0	0.4960	.	.	0.5000
PLEOMORPHIC LYMPHOMA	0	3	5	2	0.3035	0.2636	0.0217	0.1249
Pooled Lymphomas	3	6	7	5	0.3682	0.3532	0.1035	0.2358
HARDERIAN GLANDS								
ADENOCARCINOMA	0	1	0	2	0.1324	0.2636	.	0.4906
ADENOMA	6	4	4	6	0.4986	0.4258	0.5406	0.5857
Adenoma/Adenocarcinoma	6	5	4	7	0.4221	0.5567	0.5406	0.4728
JEJUNUM								
ADENOCARCINOMA	0	0	1	0	0.5000	.	0.4655	.
KIDNEYS								
HAEMANGIOSARCOMA	0	0	0	1	0.2823	0.5179	.	.
TUBULAR ADENOMA	0	2	0	3	0.1000	0.1388	.	0.2453
TUBULAR CARCINOMA	1	0	0	0	0.7500	0.5179	0.4655	0.4906
LIVER								
HAEMANGIOMA	1	1	0	1	0.4823	0.2726	0.4655	0.7453
HAEMANGIOSARCOMA	1	0	0	0	0.7500	0.5179	0.4655	0.4906
HEPATOCELLULAR ADENOMA	7	10	8	8	0.6274	0.5876	0.3505	0.2243
HEPATOCELLULAR CARCINOMA	4	3	3	2	0.7892	0.6811	0.4029	0.4369
Hemangioma/Hemangiosarcoma	2	1	0	1	0.7078	0.5402	0.7187	0.4856
Hepatocellular Adenoma/Carcinoma	10	13	11	9	0.8066	0.5648	0.2836	0.1673

Table A.3.4. (cont.) Incidence of Neoplasms in Male Mice

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
LN MESENTERIC								
HAEMANGIOMA	1	0	0	0	0.7500	0.5179	0.4655	0.4906
LUNGS & BRONCHI								
BRONCHIOLOALVEOLAR ADENOCARCINOMA	2	3	3	7	0.0541	0.1086	0.4332	0.4816
BRONCHIOLOALVEOLAR ADENOMA	8	9	6	10	0.4296	0.4577	0.4754	0.5000
Bronch. Adenoma/Adenocarc.	9	11	9	17	0.0578	0.0842	0.4397	0.3904
PANCREAS								
ISLET CELL ADENOMA	0	2	0	0	0.7520	.	.	0.2358
SARCOMA	0	1	0	0	0.5000	.	.	0.4906
PITUITARY								
ADENOMA - PARS DISTALIS	1	0	1	0	0.6423	0.5179	0.7187	0.4906
SEMINAL VESICLES								
ADENOMA	0	0	0	1	0.2823	0.5179	.	.
SKELETAL MUSCLE								
OSTEOSARCOMA	0	0	1	0	0.5040	.	0.4746	.
SARCOMA - NOS	0	0	0	1	0.2823	0.5179	.	.
SKIN (PROTOCOL)								
FIBROMA	0	0	0	1	0.2823	0.5179	.	.
FIBROSARCOMA	0	4	3	4	0.1493	0.0694	0.0948	0.0555
Fibroma/Fibrosarcoma/Sarcoma	0	4	3	6	0.0418	0.0182	0.0948	0.0555
MALIGNANT FIBROUS HISTIOCYTOMA	1	0	0	0	0.7440	0.5088	0.4576	0.4815
SARCOMA	0	0	0	1	0.2823	0.5179	.	.
SQUAMOUS CELL PAPILLOMA	1	0	1	0	0.6423	0.5179	0.7187	0.4906
TRICHOEPITHELIOMA	0	0	0	1	0.2823	0.5263	.	.
SPLEEN								
HAEMANGIOMA	0	0	0	1	0.2823	0.5179	.	.
HAEMANGIOSARCOMA	1	0	1	0	0.6423	0.5179	0.7187	0.4906
Hemangioma/Hemangiosarcoma	1	0	1	1	0.4298	0.2636	0.7187	0.4906
STOMACH								
SQUAMOUS CELL PAPILLOMA	0	0	1	0	0.5000	.	0.4655	.
Systemic								
Hemangioma	2	1	0	2	0.5357	0.3469	0.7187	0.4856
Hemangioma/Hemangiosarcoma	4	1	1	3	0.5426	0.5607	0.7775	0.8129
Hemangiosarcoma	2	0	1	1	0.5633	0.5273	0.4474	0.7453
TESTES								
INTERSTITIAL (LEYDIG) CELL ADENOMA	4	1	2	2	0.7055	0.6993	0.5956	0.8129
THORAX								
MESOTHELIOMA	1	0	0	0	0.7500	0.5179	0.4655	0.4906

Table A.3.5. Incidence of Neoplasms in Female Mice

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
ADIPOSE TISSUE								
HAEMANGIOMA	1	0	0	0	0.7460	0.4754	0.4839	0.5224
ADRENALS								
PHAEOCHROMOCYTOMA	1	0	1	0	0.5973	0.4754	0.7377	0.5224
SUBCAPSULAR CELL ADENOMA	1	0	2	1	0.3256	0.7290	0.4754	0.5224
BONE								
OSTEOMA	0	2	0	0	0.7153	.	.	0.2766
OSTEOSARCOMA	0	0	1	0	0.4683	.	0.4839	.
BRAIN								
ASTROCYTOMA	0	0	0	1	0.2302	0.4754	.	.
CAECUM								
ADENOCARCINOMA	2	0	0	0	0.9370	0.7290	0.7377	0.7757
COLON								
ADENOCARCINOMA	1	0	0	0	0.7402	0.4677	0.4762	0.5147
DUODENUM								
ADENOMA	1	1	0	2	0.2595	0.4625	0.4839	0.2766

Table A.3.5. (cont.) Incidence of Neoplasms in Female Mice

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
FEMUR								
HAEMANGIOMA	0	0	1	0	0.4683	.	0.4839	.
GALL BLADDER								
PAPILLOMA	0	1	0	0	0.4646	.	.	0.5294
H'POIETIC TUMOUR								
HISTIOCYTIC SARCOMA	2	2	1	2	0.4790	0.6430	0.4637	0.3407
IMMUNOBLASTIC LYMPHOMA	0	1	0	0	0.4683	.	.	0.5224
LYMPHOCYTIC/LYMPHOBLASTIC LYMPHOMA	7	8	7	8	0.3388	0.4727	0.6179	0.4157
MYELOID LEUKAEMIA	0	1	0	0	0.4646	.	.	0.5294
PLEOMORPHIC LYMPHOMA	4	7	2	5	0.4593	0.3964	0.6182	0.3089
Pooled Lymphomas	11	15	9	13	0.3072	0.2983	0.5744	0.4100
HARDERIAN GLANDS								
ADENOCARCINOMA	0	0	1	0	0.4683	.	0.4839	.
ADENOMA	3	2	0	2	0.6316	0.4687	0.8689	0.5567
Adenoma/Adenocarcinoma	3	2	1	2	0.5875	0.4687	0.6688	0.5567
LIVER								
HAEMANGIOMA	1	0	1	0	0.5916	0.4677	0.7296	0.5147
HEPATOCELLULAR ADENOMA	1	1	1	1	0.4912	0.7290	0.7460	0.2691
Hemangioma/Hemangiosarcoma	1	0	1	0	0.5916	0.4677	0.7296	0.5147
LN MESENTERIC								
HAEMANGIOMA	0	0	0	1	0.2302	0.4754	.	.
LUNGS & BRONCHI								
BRONCHIOLOALVEOLAR ADENOCARCINOMA	4	1	2	3	0.4169	0.4435	0.6487	0.8488
BRONCHIOLOALVEOLAR ADENOMA	8	13	14	11	0.1821	0.2390	0.0806	0.2802
Bronchioalv. Adenoma/Adenocarcoma	12	14	16	14	0.1665	0.3223	0.2277	0.4196
MAMMARY PROTOCOL								
ADENOACANTHOMA	2	1	1	1	0.5860	0.4625	0.4879	0.5341
MAMMARY ADENOCARCINOMA	2	5	4	5	0.1563	0.1751	0.3213	0.2646
MAMMARY CARCINOSARCOMA	0	1	0	1	0.2900	0.4754	.	0.5294
OVARIES+OVIDUCTS								
CYSTADENOMA	2	2	0	1	0.7327	0.4625	0.7377	0.3416
GRANULOSA CELL TUMOUR	0	0	2	0	0.4641	.	0.2381	.
LUTEOMA	0	1	0	1	0.2909	0.4754	.	0.5224
SERTOLIFORM TUBULAR ADENOMA	0	1	0	0	0.4683	.	.	0.5224
YOLK SAC CELL TUMOUR	1	0	0	0	0.7460	0.4754	0.4839	0.5224
PANCREAS								
ISLET CELL ADENOMA	1	0	0	1	0.4640	0.7290	0.4839	0.5224
PITUITARY								
ADENOMA - PARS DISTALIS	3	5	6	3	0.4848	0.6156	0.2209	0.4069
SCHWANNOMA	1	0	0	0	0.7460	0.4754	0.4839	0.5224
SKELETAL MUSCLE								
FIBROSARCOMA	0	0	1	0	0.4683	.	0.4839	.
HAEMANGIOMA	0	0	1	0	0.4683	.	0.4839	.
OSTEOSARCOMA	0	1	0	0	0.4646	.	.	0.5294
SKIN (PROTOCOL)								
FIBROSARCOMA	2	0	0	0	0.9340	0.7208	0.7296	0.7682
Fibroma/Fibrosarcoma/Sarcoma	2	0	1	0	0.8052	0.7208	0.4762	0.7682
HAEMANGIOSARCOMA	0	1	0	0	0.4646	.	.	0.5294
MALIGNANT HAIR FOLLICLE TUMOUR	1	0	0	0	0.7460	0.4754	0.4839	0.5224
MALIGNANT SCHWANNOMA	0	1	0	0	0.4646	.	.	0.5294
SARCOMA	0	0	1	0	0.4724	.	0.4921	.
SPLEEN								
HAEMANGIOSARCOMA	0	0	0	1	0.2302	0.4754	.	.
Hemangioma/Hemangiosarcoma	0	0	0	1	0.2302	0.4754	.	.
STOMACH								
SQUAMOUS CELL PAPILLOMA	0	1	0	0	0.4683	.	.	0.5224
Systemic								
Hemangioma	2	1	6	2	0.2378	0.6430	0.1092	0.5224
Hemangioma/Hemangiosarcoma	2	3	6	3	0.2263	0.4376	0.1092	0.5551
Hemangiosarcoma	0	2	0	1	0.4097	0.4754	.	0.2839

Table A.3.5. (cont.) Incidence of Neoplasms in Female Mice

Organ / Tumor	Incidence Veh	Low	Med	High	p-values Trend	Veh vs Hi	Veh vs Med	Veh vs Low
THYMUS								
THYMOMA	0	1	0	0	0.4646	.	.	0.5294
THYROIDS								
FOLLICULAR CELL ADENOMA	0	1	0	0	0.4683	.	.	0.5224
UTERINE CERVIX								
ENDOMETRIAL POLYP	0	1	2	0	0.4126	.	0.2300	0.5294
ENDOMETRIAL STROMAL CELL SARCOMA	1	0	0	0	0.7402	0.4677	0.4762	0.5147
Endo. Polyp/Strom.Cell Sarc.	1	1	2	0	0.6351	0.4677	0.4637	0.2685
FIBROMA	0	1	0	0	0.4646	.	.	0.5294
LEIOMYOMA	1	1	0	1	0.4155	0.7290	0.4839	0.2691
LEIOMYOSARCOMA	0	2	0	1	0.4126	0.4754	.	0.2766
Leiomyoma/Leiomyosarcoma	1	3	0	2	0.4566	0.4625	0.4839	0.3529
SQUAMOUS CELL CARCINOMA	1	0	0	0	0.7460	0.4754	0.4839	0.5224
UTERUS								
ENDOMETRIAL ADENOCARCINOMA	1	0	0	0	0.7460	0.4754	0.4839	0.5224
ENDOMETRIAL ADENOMA	0	0	1	0	0.4683	.	0.4839	.
ENDOMETRIAL POLYP	9	8	15	9	0.2216	0.5151	0.1183	0.5824
HAEMANGIOMA	0	1	3	1	0.2404	0.4754	0.1074	0.5224
HAEMANGIOSARCOMA	0	1	0	0	0.4646	.	.	0.5294
LEIOMYOMA	1	2	3	2	0.2952	0.4754	0.2820	0.5341
MIXED MUELLERIAN TUMOUR	0	1	0	0	0.4646	.	.	0.5294
VAGINA								
LEIOMYOMA	1	0	0	0	0.7402	0.4677	0.4762	0.5147

Appendix 4. References

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

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

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

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

Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001), *Bayesian Survival Analysis*, Springer Verlag, New York.

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 Biopharma-ceutical Statistics*. 8(1), 1-15.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325--337.

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

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

Rahman, M.A. and Lin, K.K. (2008), A Comparison of False Positive Rates of Peto and Poly-3 Methods for Long Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman, *Journal of Biopharmaceutical Statistics*, 18, 949-958. STP Peto Working Group (2002), Statistical Methods for Carcinogenicity Studies, *Toxicologic Pathology*. 30 (3), 403-414.

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

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22512	ORIG-1	BOEHRINGER INGELHEIM PHARMACEUTICALS INC	PRADAXA (DABIGATRAN ETEXILATE MESYLATE)

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
03/08/2010

KARL K LIN
03/09/2010
Concur with review