

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

22-074

STATISTICAL REVIEW(S)

8/27/07



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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 22-074
Drug Name: Somatuline Depot (lanreotide acetate) Injection 60, 90, 120 mg
Indication(s): Treatment of Acromegaly
Applicant: Beaufour Ipsen Pharma
Date(s): Received 10/27/06; user fee (10 months) 8/27/07
Review Priority: Standard

Biometrics Division: Biometrics 2
Statistical Reviewer: Lee-Ping Pian, Ph.D.
Concurring Reviewers: Todd Sahlroot, Ph.D.

Medical Division: Metabolic and Endocrine Products
Clinical Team: Eileen Craig, M.D.
Project Manager: Jennifer Johnson

Keywords: NDA review, clinical studies

Table of Contents

1. EXECUTIVE SUMMARY	3
INTRODUCTION AND BACKGROUND	3
2. DATA SOURCES	4
3. STATISTICAL EVALUATION	4
3.1 EVALUATION OF EFFICACY.....	5
3.1.1 STUDY E-28-52030-717.....	5
3.2 EVALUATION OF SAFETY DURING DOUBLE-BLIND PHASE	14
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	14
4.1 GENDER, RACE AND AGE	14
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS.....	16
5. LABELING COMMENTS:	20

1. EXECUTIVE SUMMARY

Introduction and Background

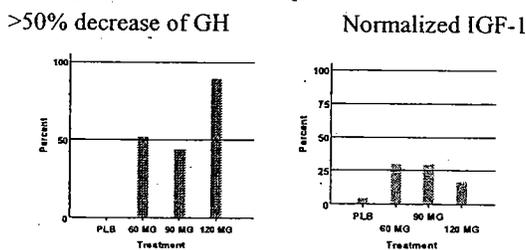
Somatuline (lanreotide acetate) Injection 60, 90, and 120 mg is a prolonged release formulation of lanreotide intended for the deep subcutaneous injection, as a treatment for patients with acromegaly.

This review concerns one phase II randomized, double-blind, placebo-controlled study E-28-52030-717 which evaluated the efficacy and safety of a single injection of lanreotide (60, 90, or 120 mg) versus placebo at Week 4 in patients with acromegaly. The primary efficacy variable was the proportion of patients with a >50% decrease in mean GH (growth hormone) from baseline to week 4. The second efficacy variable was the proportion of patients with normalized IGF-1 (insulin-like growth factor-1). Table 1 displays the analysis results of pairwise comparisons for the primary and secondary efficacy variables. For the primary efficacy variable, all 3 doses were statistically significantly better than placebo which had a 0 response rate. For normalized IGF-1, the exact test comparing lanreotide 120 mg and placebo was not significant (17% vs. 4%, p=0.2). Both the pair wise comparisons of 60 mg and 90 mg vs. placebo (30% vs. 4%) were statistically significant (p=0.02). The bar graphs display results on these endpoints (Fig. 1).

Table 1 Responder analysis results – primary and secondary efficacy variables

	Placebo n=25	Lanreotide autogel		
		60 mg n=27	90 mg n=27	120 mg n=29
GH – primary efficacy				
# of responder (%)	0 (0%)	14 (52%)	12 (44%)	26 (90%)
lanreotide minus placebo [95% CI]		52% [32, 71]	44% [25, 65]	90% [72, 98]
p-value (exact)		p<0.0001	p=0.0002	p<0.0001
IGF-1 – secondary efficacy				
# of responder (%)	1(4%)	8 (30%)	8 (30%)	5 (17%)
lanreotide minus placebo [95% CI]		26% [7, 44]	26% [7, 44]	13% [-3, 29]
p-value (exact)		p=0.02	p=0.02	p=0.2

Figure 1 Percent of responders at Week 4



In conclusion, the double-blind, placebo controlled phase of study E-28-52030-717 showed statistically significant differences between all doses of lanreotide and placebo on the primary efficacy variable, response as defined by a >50% reduction of GH from baseline.

2. DATA SOURCES

The data is extracted from the integrated analysis dataset located in the link below.

[\\Cdsesub1\nonecfd\N22074\N_000\2006-10-27\m5\datasets\cse-report](#)

3. STATISTICAL EVALUATION

Statistical Methodologies

For the 3 doses of lanreotide vs. placebo, the sponsor applied Fisher's exact test with permutation resampling for multiple comparison adjustment on the primary efficacy endpoint of percent of patients who had a >50% decrease from baseline in mean GH at week 4 of the double-blind period. The significance level of the test for each 2x2 table is calculated as the percent of such tables with a more extreme outcome than the observed result using as a denominator the results from re-sampling performed for 10,000 simulations.

Study E-28-52030-717

The study was a phase II, multi-center, randomized study conducted in patients with acromegaly who may or may not have been previously treated by surgery, radiotherapy, somatostatin analogs or dopamine agonists. Patients were randomized into one of 6 treatment groups; lanreotide 60 mg, lanreotide 90 mg, lanreotide 120 mg, placebo 60 mg, placebo 90 mg, or placebo 120 mg. The study consisted of 4 phases:

1. Wash-out phase (week -12 to week 0) for previously somatostatin analog or a dopaminergic agonist treated patients.
2. Double-blind, placebo-controlled phase (week 0 to week 4) of single injection of randomized dose of lanreotide or placebo.
3. Single-blind, fixed-dose phase (week 4 to week 20) of 4 injections of lanreotide 60, 90 or 120 mg based on dose group assigned during double-blind phase.
4. Open-label dose titration phase (week 20 to week 52) with 8 injections of lanreotide at a dose based on dose titration schema. Two dose adjustments could be made.

Patient Disposition

A total of 220 patients were screened and 111 were randomized: 27 to lanreotide 60 mg, 28 to lanreotide 90 mg, 29 to lanreotide 120 mg and 27 to placebo. Of the 111 patients randomized, 108 were injected with double-blind treatment. The three patients not injected were 1 lanreotide 90 mg patient and 2 placebo patients (Table 2).

Table 2 Patient disposition (Double-blind phase)

	PLACEBO	60 mg	90 mg	120 mg
Randomized	27	27	28	29
Injected	25	27	27	29

	PLACEBO	60 mg	90 mg	120 mg
Reason for withdrawal				
Adverse event	1 (4%)	0	0	0
Lack of efficacy	0	0	0	0
Completed double-blind	24 (96%)	27	27	29

Table 3 displays demographics by lanreotide dose groups and placebo for the randomized double-blind phase. Males and females were approximately equal except for the 90 mg lanreotide group which had 2/3 of females.

Table 3 Patient demographics

LANREOTIDE				
	60 mg n=27	90 mg n=27	120 mg n=29	Placebo n=25
Age (years)				
Mean (SD)	52.2 (16.6)	54.5 (14.2)	55.6 (12.1)	50.5 (12.1)
Min, Max	19,84	27,77	24,78	27,72
Gender				
M	13 (48%)	9 (33%)	16 (55%)	13 (52%)
F	14 (52%)	18 (67%)	13 (45%)	12 (48%)
Race				
CAUCASIAN	24 (89%)	24 (89%)	25 (86%)	20 (80%)
ASIAN, ORIENTAL	2 (7%)	2 (7%)	3 (10%)	4 (16%)
BLACK	1 (4%)	1 (4%)	1 (3%)	1 (4%)

3.1 Evaluation of Efficacy

3.1.1 Study E-28-52030-717

Table 4 displays descriptive statistics for GH in the double-blind period (4 weeks). In general median and mean values are not similar because of outliers (Figs 2 and 3). The percent change from baseline in the placebo group had the greatest standard deviation (171) and range (865). The maximum GH percent change was +834% (4.3 to 40 ng/mL).

Table 4 Descriptive statistics of GH – Double-blind (Week 4)

Double-blind period	Lanreotide autogel			Placebo n=24
	60 mg n=27	90 mg n=27	120 mg n=29	
Baseline				
Median	12	10	9	9
Mean (SD)	25 (46)	17 (18)	19 (20)	19 (28)
[min, max]	[3, 244]	[3, 67]	[3, 83]	[3, 131]
Range	241	64	79	128
Week 4				
Median	6	5	3	13
Mean (SD)	12 (21)	9 (10)	4 (3)	23 (29)
[min, max]	[1, 107]	[1, 44]	[1, 17]	[3, 131]
Range	107	40	16	128
Change from baseline				
Median	-7	-4	-7	+1

Double-blind period	Lanreotide autogel			
	60 mg n=27	90 mg n=27	120 mg n=29	Placebo n=24
Mean (SD)	-13 (26)	-9 (12)	-15 (18)	3.9 (8)
[min, max]	[-137, 1]	[-49, 5]	[-8, 0]	[-4, 36]
Range	138	54	81	40
% change from baseline				
Median	-56%	-40%	-75%	+9%
Mean (SD)	-45% (31)	-41% (39)	-70% (22)	55% (171)
[min, max]	[-93, 15]	[-95, 94]	[-98, -2]	[-31, 834]
Range	108	189	96	865

Figure 2 Box plot of GH at baseline and Week 4

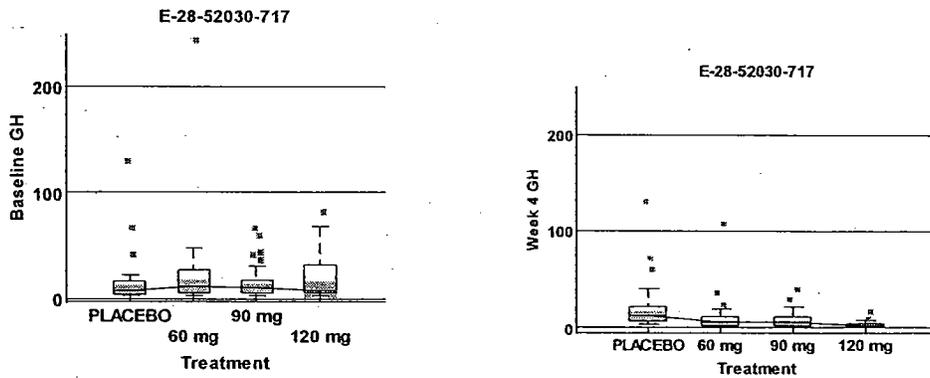


Figure 3 Box plot of GH change and % change from baseline

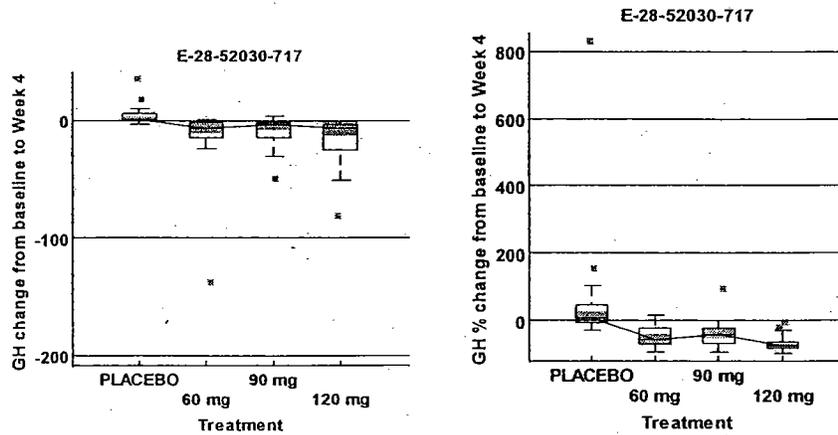


Figure 4 Cumulative distribution of GH change and % change from baseline

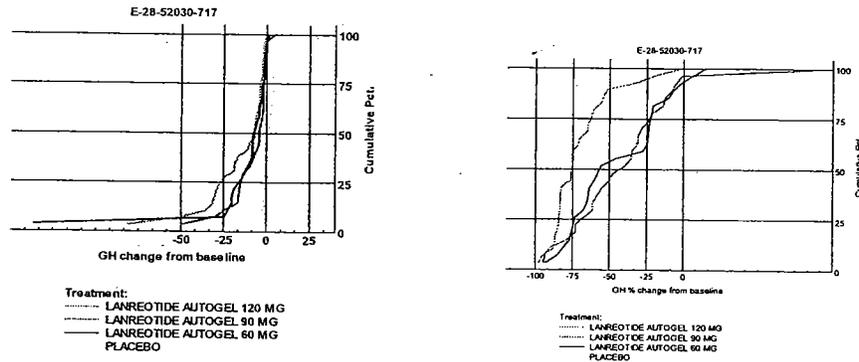


Table 5 displays median GH percent change by the 6 randomization groups over the study period. After the double-blind 1st injection, patients on placebo were administered lanreotide with their randomized doses for injections 2 to 5 (single-blind). GH value at week 4 prior to lanreotide injection was used as baseline for those placebo/lanreotide groups.

At Week 16, the 3 placebo-lanreotide treated groups were similar to the 3 lanreotide-treated groups in the median GH percent reduction which was approximately 80% in the 120 mg groups while the other 2 fixed doses were $\geq 70\%$

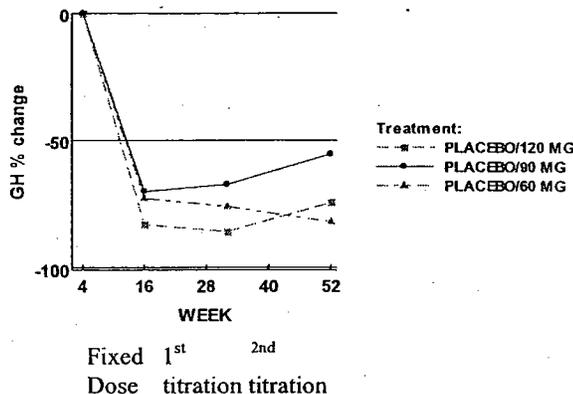
After the 1st titration (injections 6 to 9) and 2nd titration (injections 10 to 13), the lanreotide-treated groups were similar in the median GH % change from baseline (in the high 70%).

Table 5 Median GH % change by randomization groups over time

Week evaluation	Placebo(week 4)/Lanreotide			Lanreotide		
	60 mg n=7	90 mg n=9	120 mg n=8	60 mg n=27	90 mg n=27	120 mg n=29
4	+10	+9	+16	-56	-40	-75
double blind						
13	-82	-66	-76	-79	-74	-85
14	-79	-66	-80	-76	-75	-82
15	-74	-70	-85	-76	-72	-83
16	-73	-70	-83	-68	-70	-84
fixed dose						
single blind						
32	-76	-67	-86	-77	-80	-81
1st titration						
52	-82	-56	-75	-77	-78	-77
2nd titration						

Figure 5 Median GH % change by placebo/lanreotide groups – Fixed-dose and dose titrations

STUDYID
E-28-52030-717



Primary Efficacy endpoint:

The primary efficacy variable is the proportion of patients with a >50% decrease in mean GH from baseline at week 4 after a single injection. The percent of patients with >50% response was 0 in the placebo-treated patients. The Fisher Exact test results showed all 3 doses of lanreotide were statistically significantly better than placebo (Table 6). The 120 mg lanreotide dose had the highest response rate (90%) while the 90 mg and 60 mg doses had response rates of 44% and 52%, respectively.

In addition to the categorical analysis GH change and percent change were also analyzed using ANCOVA (analysis of covariance). Similar to the responder analysis the analysis of covariance results showed all doses were statistically significantly better than placebo (Table 6, Figure 6).

Table 6 Responder analysis results

	Lanreotide autogel			
	60 mg	90 mg	120 mg	Placebo
# Responder/treatment n	14/27	12/27	26/29	0/25
Difference from placebo [95% CI]	52% [32%, 71%]	44% [25%, 65%]	90% [72%, 98%]	
p value of drug vs. placebo	<0.0001	0.0002	<0.0001	

Figure 6 Difference of proportions between lanreotide and placebo in GH responders (>50% reduction)

APPEARS THIS WAY
ON ORIGINAL

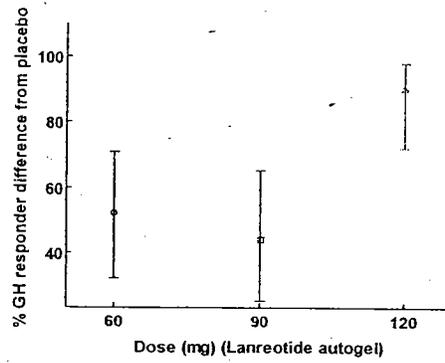


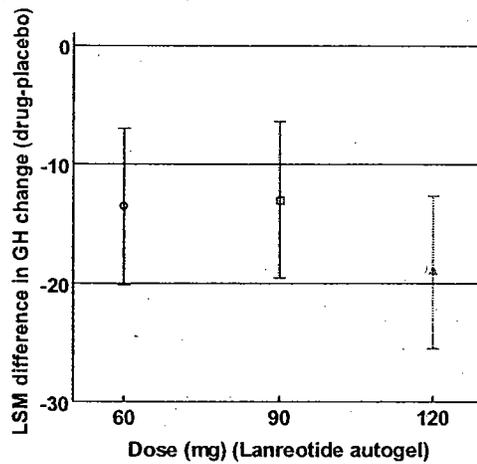
Table 7 Analyses of GH % change from baseline – double-blind (Week 4)

Double-blind period	Lanreotide autogel			Placebo n=24
	60 mg n=27	90 mg n=27	120 mg n=29	
Week 4 LSM (SE)	-43.4 (16.5)	-41.6 (16.4)	-70.5 (15.9)	55 (17.4)
LSM Difference (SE) [95% CI] lanreotide - placebo	-98.3 (24) [-146, -50.7]	-96.6 (24) [-144.1, -49.1]	-125.5 (23.6) [-172.2, -78.7]	
p-value	<0.0001	0.0001	<0.0001	
Median	-56	-40	-75	9.4
Median 2-sample test				
overall median (n)	-14 (n=51)	-14 (n=51)	-31 (n=53)	
# of points above median (expected under Ho)	5 vs. 20 (13 vs. 12)	5 vs. 20 (13 vs. 12)	2 vs. 24 (14 vs. 12)	
	p<0.0001	p<0.0001	p<0.0001	

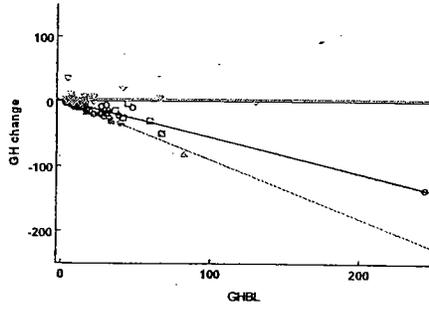
Table 8 Analysis of GH change from baseline – double-blind (Week 4)

Double-blind period	Lanreotide autogel			Placebo n=24
	60 mg n=27	90 mg n=27	120 mg n=29	
Week 4 LSM (SE)	-10.4 (1.9)	-9.7 (1.9)	-15.8 (1.8)	3.2 (2)
Lanreotide - placebo LSM Difference (SE) [95% CI]	-13.6 (2.8) [-20.2, -7]	-12.9 (2.8) [-19.5, -6.3]	-19 (2.7) [-25.5, -12.6]	
p-value	<0.0001	0.0001	<0.0001	
Median test Overall Median		-1.3 p=0.000004	-1.1 p=0.000482	
Median	-6.4	-3.7	-6.5	+1.3
Median 2-sample test overall median (n) # of points above median lanreotide vs. placebo (expected under Ho)	-1.1 (n=51) 5 vs. 20 (13 vs. 12)	-1.3 (n=51) 5 vs. 20 (13 vs. 12)	-2.2 (n=53) 2 vs. 24 (14 vs. 12)	
	p=0.0007	p<0.0001	p<0.0001	

Figure 7 LSM difference between doses of Lanreotide and placebo (95% CI) in GH change

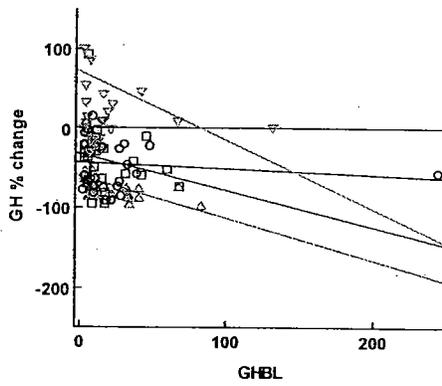


E-28-52030-717



Treatment:
—△— LANREOTIDE AUTOGEL 120 MG
—□— LANREOTIDE AUTOGEL 90 MG
—○— LANREOTIDE AUTOGEL 60 MG
—▽— PLACEBO

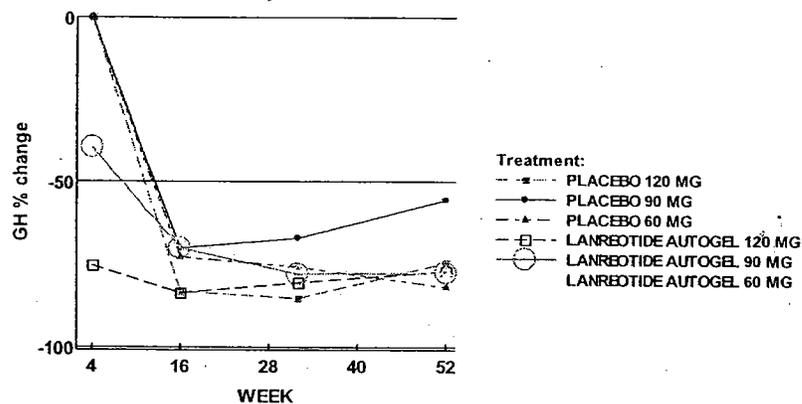
E-28-52030-717



Treatment:
—△— LANREOTIDE AUTOGEL 120 MG
—□— LANREOTIDE AUTOGEL 90 MG
—○— LANREOTIDE AUTOGEL 60 MG
—▽— PLACEBO

Figure 8 Median GH % change for weeks 4, 16, 32 and 52 by randomization group

STUDYID
E-28-52030-717



At baseline, 52% (56/108) of the patients were naïve or had stopped treatment for at least 3 months prior to study entry. The mean GH inclusion value was >5 ng/mL for these 52% patients at visit one. The GH inclusion criteria for other patients who were on treatment (somatostatin analog or dopamine agonist) prior to study entry was GH>3 ng/mL and at least a 100% increase in mean GH after washout of medication.

For exploratory analysis prior treatment status was included as a factor in the ANCOVA model even though it is not a randomization stratum. The treatment-by-prior treatment interaction of GH percent change was significant ($p=0.0045$) for lanreotide vs. placebo. When excluding outliers of >800% GH percent change, the interaction was more significant due to the decrease of standard error ($p=0.0028$). Table 9 displays the descriptive statistics by the prior-treatment strata. The LSM differences of the 2 prior treatment strata in Table 10 showed that more GH percent change difference between lanreotide and placebo in the previous-treated patients than naïve patients. The GH in the placebo group of prior treated stratum increased from baseline for all but one (no change) patients while GH in 70% of the naïve placebo patients decreased from baseline. The interaction is driven by the difference in placebo response in the 2 strata who were entered in the study based on different criteria.

The box plot (Fig 9) displays the median, outliers and 95% confidence interval for the median. The 95% confidence interval of the median was narrowest and the standard deviation of the mean (15%) the smallest for the 120 mg lanreotide treatment among the 4 treatment groups.

Figure 9 Box plot of GH % change from baseline by prior medication

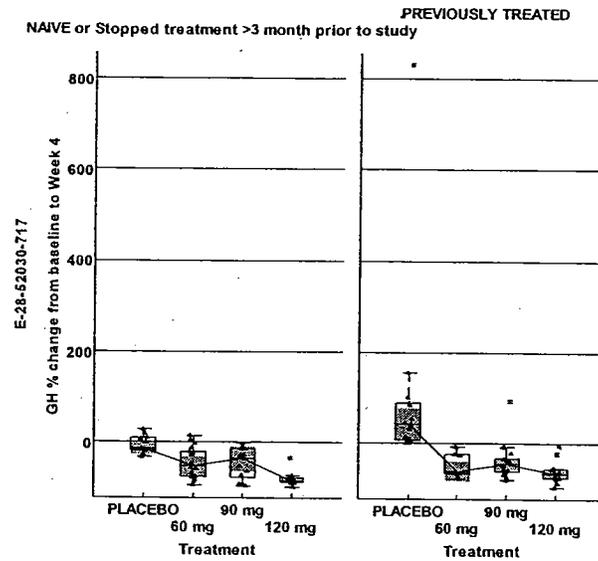


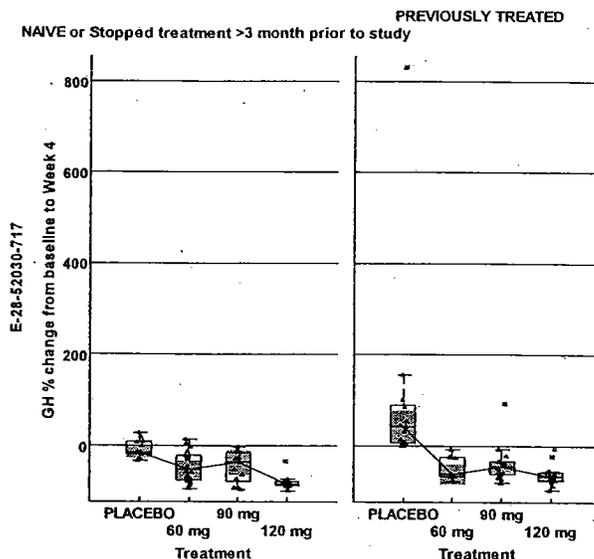
Table 9 Descriptive statistics of GH % change from baseline by prior-medication

	Naïve or stopped treatment >3 months				Prior treatment			
	Placebo n=11	60 mg n=18	90 mg n=13	120 mg n=14	Placebo n=13	60 mg n=9	90 mg n=14	120 mg n=15
Median	-13	-51	-35	-84	42	-61	-44	-65
Mean(SD)	-5(20)	-45(34)	-45(34)	-80(15)	106(223)	-46(26)	-37(43)	-61(24)
[min, max]	[-31, 30]	[-93, 15]	[-94, -1]	[-98, -31]	[0, 834]	[-77, -7]	[-80, 94]	[-95, -2]
Range	61	108	93	67	834	70	174	94

Table 10 ANCOVA* of GH % change from baseline by prior-medication strata

	Naïve or stopped treatment >3 months				Prior treatment			
	Placebo n=11	60 mg n=18	90 mg n=13	120 mg n=14	Placebo n=13	60 mg n=9	90 mg n=14	120 mg n=15
LSM (SE)	-7 (25)	-39 (20)	-44 (23)	-78 (22)	107 (23)	-51 (28)	-40 (22)	-64 (21)
[95% CI]	[-56, 42]	[-78, 0]	[-90, 1]	[-122, -34]	[62, 153]	[-106, 4]	[-83, 4]	[-106, -21]
LSM difference from placebo (SE) [95% CI]		-32 (45)	-32 (48)	-65 (45)		-158 (51)	-153 (45)	-176 (42)
		[-123, 59]	[-129, 64]	*[-156, 26]		[-260, -56]	[-244, -62]	[-261, -90]

*ANCOVA model: treatment, prior-treatment strata as fixed effects and baseline GH as covariate



3.2 Evaluation of Safety during double-blind phase

Table 11 adds p values to the sponsor's table of treatment-emergent AE with incidence of rates 5% or more.

Table 11 Most Commonly ($\geq 5\%$) Reported TEAE – Double-blind phase

	PLACEBO n=25	60 mg n=27	90 mg n=27	120 mg n=29
Diarrhoea	0	3 (11%) p=0.13	10 (37%) p=0.0007	13 (45%) p=0.00004
Abdominal pain	1 (4%)	2 (7%) p=1	2 (7%) p=1	2 (7%) p=1
Bradycardia	0	3 (11%) p=0.13	2 (7%) p=1	2 (7%) p=1
Weight decrease	0	2 (7%) p=1	4 (15%) p=0.1	1 (3%) p=1
Anaemia	0	1 (4%) p=1	4 (15%) p=0.1	1 (3%) p=1
Flatulence	0	0 p=1	2 (7%) p=1	3 (10%) p=0.2

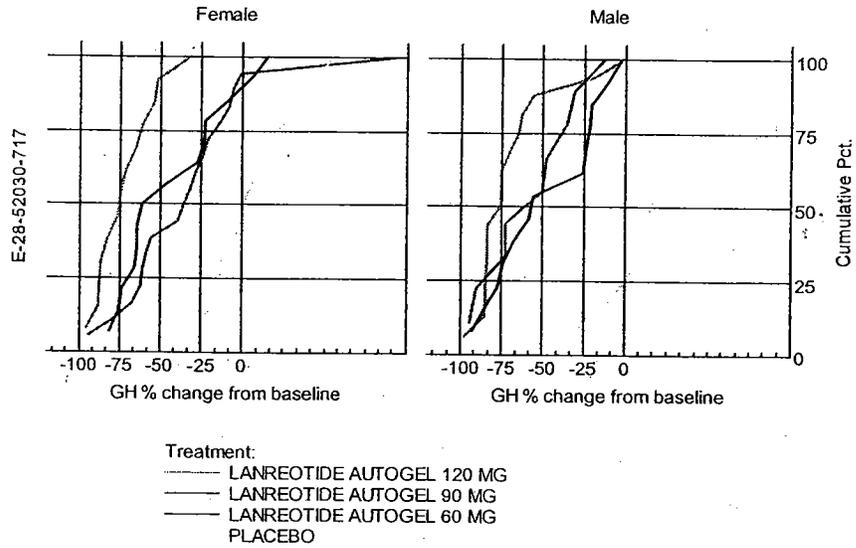
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

There were no treatment-by-subgroup interactions detected for gender, race or age group. The cumulative distributions of these subgroups are presented.

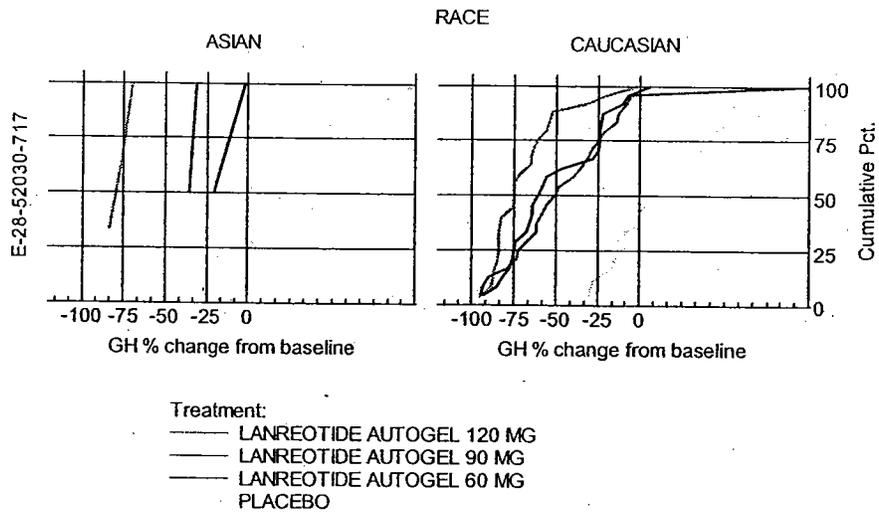
Gender:

Figure 10 % GH change from baseline by gender – double-blind phase



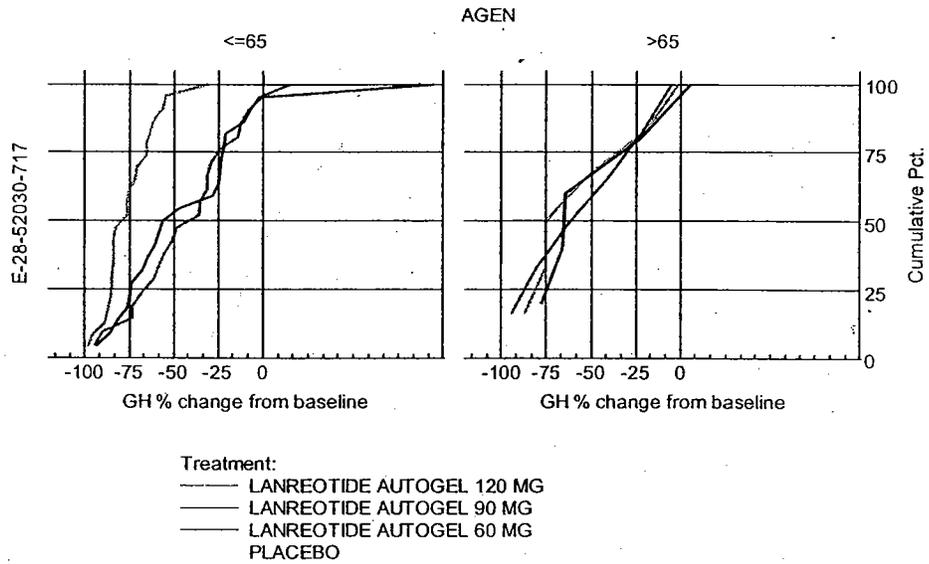
Race:

Figure 11 % GH change from baseline by race – double-blind phase



Age group:

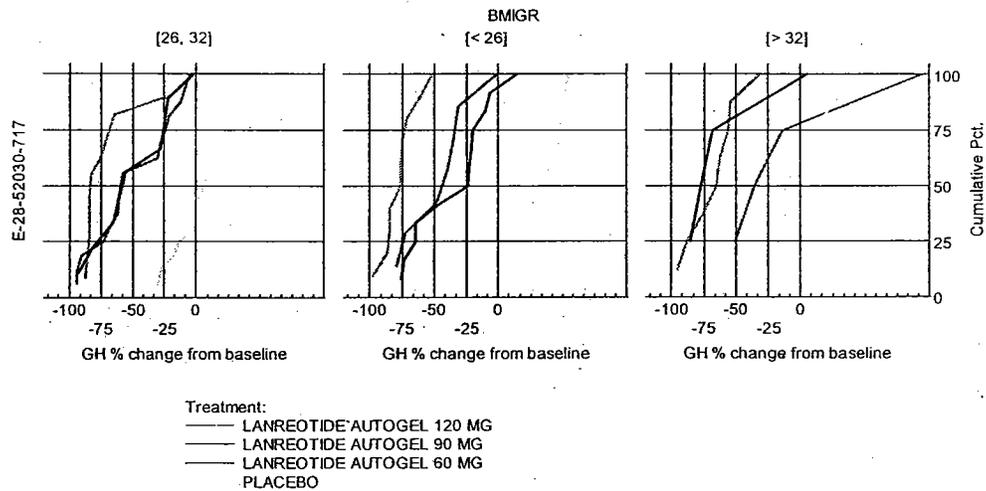
Figure 12 % GH change from baseline by age group – double-blind phase



4.2 Other Special/Subgroup Populations

BMI group:

Figure 13 % GH change from baseline by baseline BMI group – double-blind phase



Region:

Figure 14 % GH change from baseline by geographical region – double-blind phase

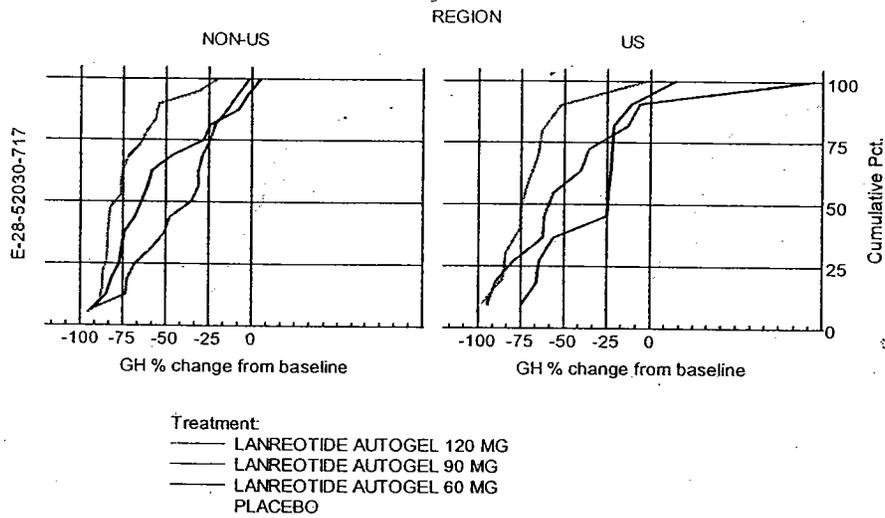
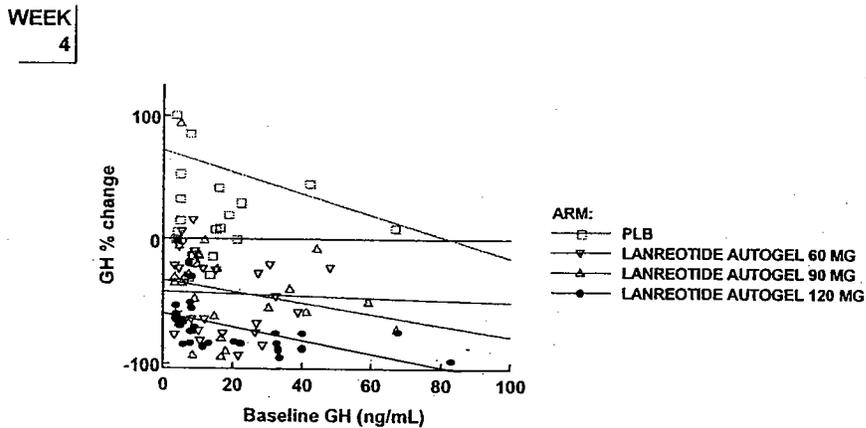
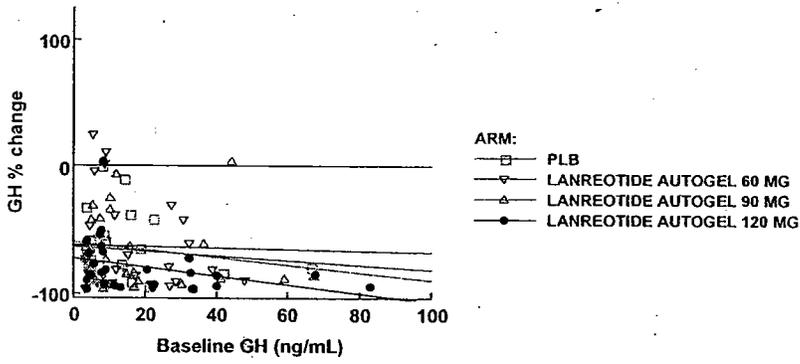


Figure 15 displays regressions of GH percent change by baseline GH over time. At Week 52, median GH % changes were improved compared to Week 4. Table 12 shows descriptive statistics for GH% change over time. The placebo data shown at time points after week 4 represent patients taking active treatment who were original randomized to placebo. Results for weeks 32 and 52 represent the effects of the 1st and 2nd dose titrations, respectively.

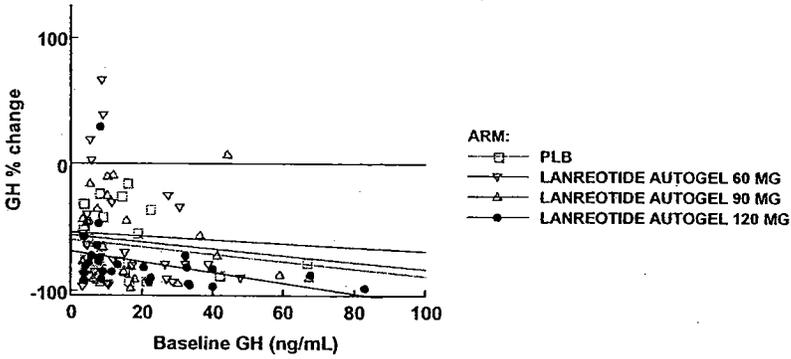
Figure 15 Regression of GH% change by baseline GH



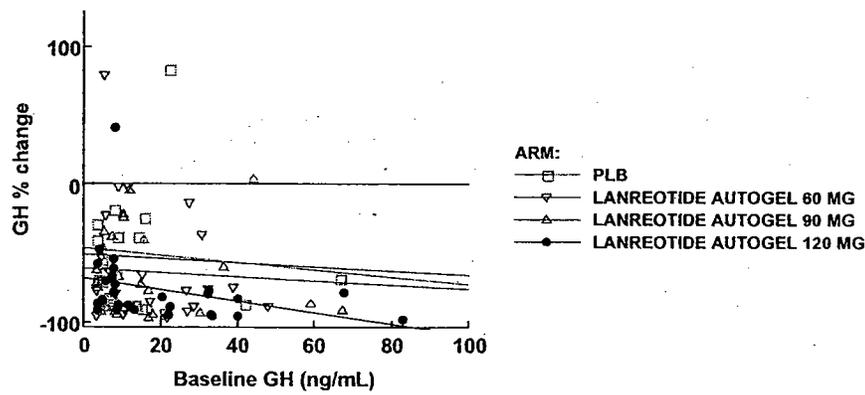
WEEK
13



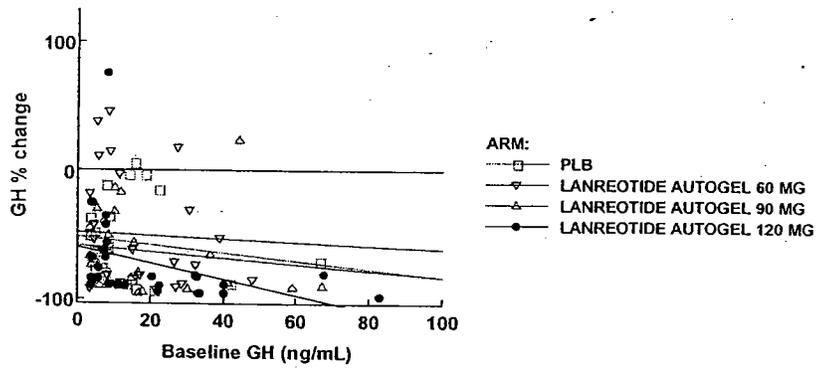
WEEK
14



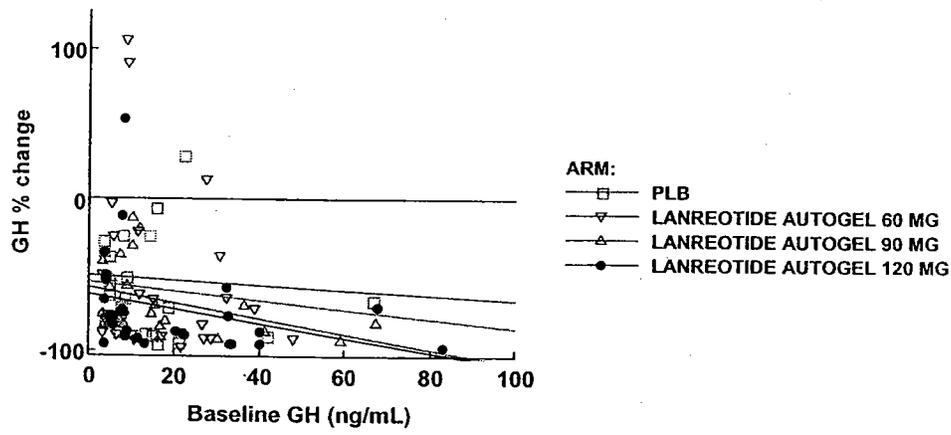
WEEK
15



WEEK
16



WEEK
32



WEEK
52

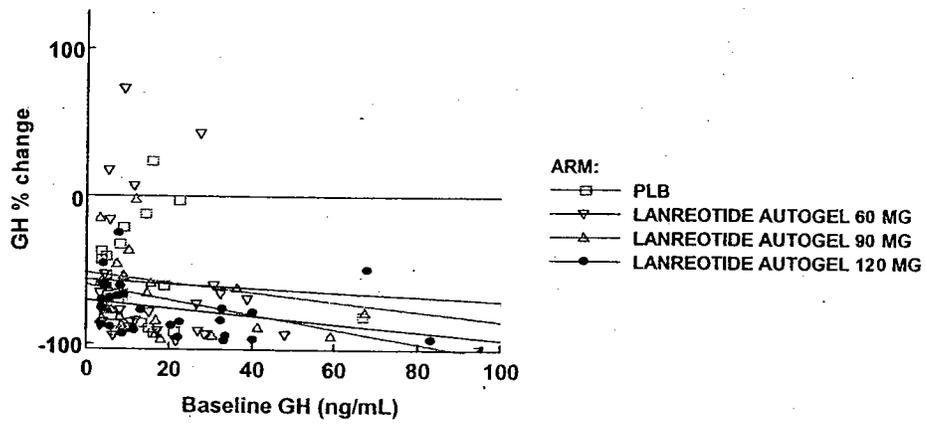


Table 12 GH descriptive statistics over time

WEEK	measure	Placebo n=24	60 mg n=27	90 mg n=27	120 mg n=29
4	Median	9	-56	-40	-75
	Mean(SD)	55 (171)	-45 (31)	-41 (39)	-70 (22)
	[min, max]	[-31, 834]	[-93, 15]	[-94, 94]	[-98, -2]
	range	865	108	188	96
13	Median	-75	-79	-74	-85
	Mean(SD)	-65 (26)	-63 (35)	-67 (29)	-79 (22)
	[min, max]	[-98, 1]	[-96, 24]	[-97, 4]	[-97, 4]
	range	98	120	101	101
14	Median	-74	-76	-75	-82
	Mean(SD)	-65 (25)	-58 (44)	-61 (31)	-78 (24)
	[min, max]	[-97, -15]	[-96, 66]	[-98, 7]	[-98, 29]
	range	82	162	104	126
15	Median	-68	-76	-72	-83
	Mean(SD)	-51 (56)	-55 (56)	-63 (31)	-76 (27)
	[min, max]	[-98, 136]	[-97, 145]	[-96, 3]	[-98, 41]
	range	234	243	100	139
16	Median	-70	-68	-70	-84
	Mean(SD)	-58 (32)	-51 (44)	-63 (31)	-71 (35)
	[min, max]	[-97, 7]	[-93, 45]	[-96, 24]	[-98, 76]
	range	104	139	119	174
32	Median	-69	-77	-78	-81
	Mean(SD)	-61 (33)	-55 (53)	-68 (24)	-73 (32)
	[min, max]	[-98, 29]	[-98, 105]	[-93, -13]	[-98, 54]
	range	127	203	80	152
52	Median	-65	-77	-78	-77
	Mean(SD)	-58 (33)	-60 (46)	-69 (25)	-76 (18)
	[min, max]	[-97, 24]	[-98, 73]	[-96, -2]	[-99, -25]
	range	121	170	94	73

5. LABELING COMMENTS:

- The proposed indication in form FDA 356h was treatment of acromegaly. However, the proposed indications and usage in the label were 1. the long-term treatment of acromegalic patients who have had an inadequate response to or cannot be treated with surgery and/or radiotherapy and 2.

For the first indication, the controlled portion of the study was only 4 weeks.

indication, therefore, is not justified.

s. The 2nd

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

/s/

Lee-Ping Pian
8/24/2007 04:48:47 PM
BIOMETRICS

Todd Sahlroot
8/27/2007 09:12:40 AM
BIOMETRICS

Thomas Permutt
8/27/2007 11:47:51 AM
BIOMETRICS
concur

6/19/07



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: 22,074 / Serial 000

Drug Name: Somatuline® Autogel (also referred to as BIM23014) [Lanreotide Acetate Injection 60/90/120 mg]

Indication:

1. long-term treatment of acromegalic patients who have had an inadequate response to or cannot be treated with surgery and/or radiotherapy.
2. _____

Applicant: Beaufour-IPSEN Pharma

Date: Submitted 10/27/2006

Review Priority: Standard

Biometrics Division: Division 6

Statistical Reviewer: Steve Thomson

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

Medical Division: Metabolism and Endocrinology Products

Toxicologist: Reviewer: Da Lin (Dylan) Yao, Ph.D.
Team Leader: Karen Davis Bruno, Ph.D.

Project Manager: Jennifer Johnson

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

Table of Contents

1. EXECUTIVE SUMMARY	3
1.1. CONCLUSIONS AND RECOMMENDATIONS	3
1.2. BRIEF OVERVIEW OF THE STUDIES	6
1.3. STATISTICAL ISSUES AND FINDINGS	6
1.3.1. <i>Statistical Issues</i>	6
1.3.2. <i>Statistical Findings</i>	9
2. INTRODUCTION	9
2.1. OVERVIEW	9
2.2. DATA SOURCES	9
3. STATISTICAL EVALUATION	9
3.1. EVALUATION OF EFFICACY	9
3.2. EVALUATION OF SAFETY	9
3.2.1. <i>Project 77006: A 104-Week Subcutaneous Injection Carcinogenicity Study in the Albino Mouse</i>	10
3.2.2. <i>Project 77005: A 104-Week Subcutaneous Injection Carcinogenicity Study of BIM 23014 in the Albino Rat</i>	16
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	21
5. SUMMARY AND CONCLUSIONS	21
5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	21
5.2. CONCLUSIONS AND RECOMMENDATIONS	21
APPENDICES:	22
APPENDIX 1. SURVIVAL ANALYSIS	22
APPENDIX 2. BAYESIAN ANALYSIS OF SURVIVAL	25
APPENDIX 3. FDA TUMORIGENICITY ANALYSIS	29
APPENDIX 4. REFERENCES	37

1. EXECUTIVE SUMMARY

According to the reports provided by the Contract Research Organization, this submission was intended to assess the carcinogenic potential of daily subcutaneous injection of compound BIM23014, i.e., Somatuline® Autogel, when administered for periods of up to 24 months in mice and rats. The sponsor was the Beaufor-IPSEN Group, in Paris, France. The studies were conducted by the _____ The descriptions of the studies below are taken from the corresponding Final Reports.

1.1. Conclusions and Recommendations

The submission summarizes the results of both a mouse study and a rat study of the carcinogenic potential of BIM23014 following daily injection for two years. In the mouse study there were seven treatment groups per gender, including two supposedly identical vehicle controls, and five treatment groups with dose levels of 0.5, 1.5, 5, 10, and 30 mg/kg/day. The latter five BIM23014 dose groups were labeled as Low, Medium, Medium-high, High, and Max dose groups, respectively. In the rat study there were five treatment groups per gender, including two identical vehicle controls, and three treatment groups with dose levels of 0.1, 0.2, and 0.5 mg/kg/day. The latter three treatment groups were labeled as Low, Medium, and High dose groups, respectively. The putatively identical vehicle control groups in each species each had 60 animals per group, while the remaining BIM23014 treatment groups had 70 animals per group. In mice there were an additional 63 toxicokinetic animals at each of the five BIM23014 doses, while in rats there were an additional 15 toxicokinetic animals at each of the three BIM23014 doses. Each animal was given a daily subcutaneous injection of the vehicle or the test article at predetermined sites (6 sites for mice and 7 for rats) where injection sites were rotated according to a fixed schedule after each injection.

The statistical significances of the tests of differences in survival across treatment groups are given below. Since differences between the two vehicle controls should be solely due to randomization, for the tests below these two control groups are pooled to a single control group. The test for homogeneity is a test that survival is equal across treatment groups, while the test of trend is a test of dose related trend. The Cox test is usually called the logrank test, while the K-W, i.e., Kruskal-Wallis test, is more commonly called the Wilcoxon test or the generalized Wilcoxon test. Note that the Wilcoxon test places more weight on earlier events than does the logrank test.

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

	Mice				Rat			
	Males		Females		Males		Females	
	Cox	K-W	Cox	K-W	Cox	K-W	Cox	K-W
Homogeneity	<0.0001	<0.0001	0.0104	0.0062	0.0226	0.0141	0.0082	0.0231
Trend over all groups	<0.0001	<0.0001	0.0004	0.0003	0.1187	0.0634	0.0080	0.0127
Departure from trend	<0.0001	0.0009	0.5507	0.4535	0.0300	0.0288	0.0824	0.1628

For both species and genders the hypotheses of homogeneity in survival is always rejected (all eight $p \leq 0.0231$). For both genders in mice and in female rats the hypotheses of no trend in mortality is also rejected (all six $p \leq 0.0127$), i.e., suggesting that there is a trend. For male rats the hypothesis of trend was arguably close to statistical significance (Cox $p=0.1187$ and K-W $p=0.0634$). In male rats and especially male mice the tests of lack of homogeneity above and beyond that explained by trend over groups were also statistically significant (all four $p \leq 0.0300$). In female mice, the tests of such departure from trend were quite statistically nonsignificant (both $p \geq 0.4535$), but somewhat close to significance in female rats (Cox $p=0.0824$ and K-W $p=0.1628$). From the incidence tables (tables 6, 7, 13, and 14 below) or the Kaplan-Meier curves in Appendix 1, one can see that in female mice, male rats, and especially male mice the highest dose group has the lowest survival, with the remaining dose groups relatively closely intertwined. Although not shown in the table above, in mice, after deleting the maximum dose group, no tests for lack of homogeneity in survival were statistically significant (all $p \geq 0.4737$). Absence of proof is not proof of absence, but this, and the results on trend, are consistent with the notion of homogeneity in survival among the remaining dose groups and controls. So, in mice, this does seem to be consistent with the observed tendency for differences in survival to be mostly due to the difference between the maximum dose group (Group 7) and the remaining dose groups. For female rats, the dose related trend is actually negative in that the control groups seem to have the lowest survival, with the three BIM23014 groups relatively intertwined but having generally higher survival. Again, further details are presented in Appendix 1. In female rats the higher survival in the BIM23014 treatment groups may be associated with lower animal weights. Note the Sponsor's assessments were similar, but not exactly the same as this reviewer's analyses (please see Sections 3.2.1.1 and 3.2.2.1). Results from an experimental Bayesian analysis of mortality are summarized in Appendix 2.

In both species, the statistically significant neoplasms were primarily at injection sites. In a discussion of these results, the toxicologist expressed the opinion that these statistically significant tests on neoplasms were not strictly evidence of drug related carcinogenicity, but were rather due to local irritation effects from the repeated injections. The endpoint used in the FDA analyses of tumorigenicity is the minimum of the time of observation, time of death due to the tumor, or time of detection when the animal dies or is sacrificed. The Sponsor's analyses of tumorigenicity are apparently based only on the later two. This should have little to no effect on actual tumor incidence, but could explain differences in the actual tests of tumorigenicity.

To adjust for the multiplicity of comparisons involved in a tumorigenicity analysis for standard rodent models, the Agency analysis followed the Haseman-Lin-Rahman rules described in Section 1.3.1.4 below. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of trend, rare tumors (background incidence <1%) should be tested at a 0.025 (2.5%) significance level and common tumors at a 0.005 (0.5%) level. Tests of pairwise differences between controls and the highest dose should be tested at a 0.05 (5%) level for rare tumors and at a 0.01 (1%) level for common tumors. For both mice and rats, potentially statistically significant tumor incidences are summarized in tables 8, 9, 15, and 16, below, with more complete incidence tables in Appendix 3.

In mice, all the nominally statistically significant tumors in tables 8 and 9 would be classified as rare tumors. At each injection site and pooled over injection sites, the tests of trend in fibrosarcoma and, except for the left lumbar site in female mice, all pairwise comparisons between the maximum dose group and the pooled controls were statistically significant (all trend $p \leq 0.0122$, less than 0.025, and all relevant pairwise comparisons $p \leq 0.0247$, less than 0.05). In male mice, at the left and right dorsal thoracic sites and the left lumbar site, tests of trend in malignant fibrous histiocytoma were all statistically significant (all three $p \leq 0.0192$). The tests of pairwise differences in malignant fibrous histiocytoma between the maximum dose group and the pooled controls in male mice was statistically significant at all pooled injection sites and at the left dorsal thoracic site ($p = 0.0025$ and $p = 0.0226$, respectively). In addition, deleting the maximum dose group, the test in female mice comparing all neoplasms over all injection sites at the high dose group (Group 6) to the pooled controls was statistically significant ($p = 0.0212$). Note however, that strictly speaking, the usual Haseman-Lin-Rahman rules do not apply to this comparison, and it can not necessarily be considered as adjusted for multiplicity.

In rats, of the potentially statistically significant differences, due to the incidence in the pooled controls, for each gender only the grouping of "any neoplasm" at any injection site would be classified as a common tumor. The remaining neoplasms cited in tables 15 and 16 below would be classified as rare tumors. Using the Haseman-Lin-Rahman rules, in male rats both the test of trend and the test of pairwise differences between the high dose group and the pooled controls in malignant lymphoma in hemolym tissue were statistically significant (trend $p = 0.0146$ and pairwise $p = 0.0389$). Further, in male rats over all injection sites, as well as the left dorsal thoracic site, and both the left and right lumbar injection sites, the tests of trend in malignant fibrous histiocytoma were all statistically significant (all three trend $p \leq 0.0217$). The corresponding pairwise comparisons at both the left and right lumbar injection sites in male rats were also statistically significant (both $p \leq 0.0390$). In the male rats, for fibrosarcoma at the left dorsal thoracic site, the test of overall trend and pairwise differences between the high dose group and the pooled controls were both statistically significant (trend $p = 0.0061$ and pairwise $p = 0.0356$). Pooling all injection sites in male rats, the tests for overall trend and pairwise differences between the high dose group and the pooled controls in fibrosarcoma were both statistically significant (trend $p = 0.0010$ and pairwise $p = 0.0161$). Pooling all neoplasms over all injection sites the tests for overall trend and pairwise differences between the high dose group and the pooled controls were both statistically significant (trend $p < 0.0001$ and pairwise $p < 0.0001$). In female rats, pooling over all injection sites, the tests of trend in fibrosarcoma, malignant fibrous histiocytoma, and all neoplasms were all statistically significant (all three $p \leq 0.0023$ – but recall that the grouping of all neoplasms is not considered a rare tumor). The tests of pairwise differences in female rats in fibrosarcoma, malignant fibrous histiocytoma, and all neoplasms were all statistically significant ($p = 0.0136$, $p = 0.0210$, and $p = 0.0020$, respectively).

1.2. Brief Overview of the Studies

One mouse study and one rat study were submitted:

Study Report 77005: A 104-Week Subcutaneous Injection Carcinogenicity Study of BIM23014 in the Albino Rat,

and

Study Report 77006: A 104-Week Subcutaneous Injection Carcinogenicity Study in the Albino Mouse.

These studies were designed to assess the carcinogenic potential of BIM23014, when administered by subcutaneous injection at dose levels of 0, 0.5, 1.5, 5, 10, or 30 mg/kg/day in mice and at dose levels of 0, 0.1, 0.2, or 0.5 mg/kg/day in rats. The Sponsor indicates that

CD-1® mice were randomized into seven study groups in mice and five study groups in rats, each with 60 animals per gender in two nominally identical vehicle controls and 70 animals in each of the remaining BIM23014 study groups. In the Sponsor's analysis these were labeled as dose groups 1-7 in mice and 1-5 in rats, where dose groups 1-2 were putatively identical controls in each study. In mice, for the FDA analysis, the five BIM23014 treatment groups were labeled as Low, Medium, Medium-high, High, and Max. In rats the three BIM23014 groups were labeled as Low, Medium, and High. In addition, there were a number of toxicokinetic animals in both studies.

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 Control Groups:

The Sponsor provides tests of the differences in mortality and tumorigenicity between the two controls. But supposedly these control groups are identical and if differences are observed the analyst is faced with a conundrum. Either a rare event occurred and differences are due solely to the results of the randomization or there were significant unplanned secular trends that render the results of the entire study questionable. The latter circumstance should be apparent by other means. Under such circumstances this reviewer does not agree that one should "waste" the probability of a wrong decision (i.e., type I error, "alpha") on such tests, and does not include tests of differences between the controls in the FDA analysis.

1.3.1.2 Survival Analysis:

Both the Cox logrank and Kruskal-Wallis-Wilcoxon tests were used to test homogeneity of survival among the treatment groups. Tests of dose related trend using a Cox proportional odds model were also performed. 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 Sponsors analyses are summarized in Sections 3.2.1.1 and 3.2.2.1.

Note that due to the severity of skin related reactions, particularly at the injection sites, in male mice the Max dose group (Group 7) males were terminated during Week 87. Max dose group female mice were terminated during Week 98. Other dose groups were terminated at or after week 104.

1.3.1.3. Tests on Neoplasms:

The FDA tumorigenicity analyses of fatal tumors are based on the time of death, and for observable tumors based on time of detection. Both are analyzed at the time of detection with an analysis equivalent to the death rate method. Non-fatal tumors found at the time of the animals' death are labeled as incidental, and were analyzed by the so-called prevalence method. For the FDA analyses all three results were pooled. The Sponsor's analyses are based solely on fatal and incidental tumors. The tests on these neoplasms used in the FDA analysis are basically tests of trend. For the mice, significance levels of three tests are provided: 1) a test of trend over from the pooled controls over the five BIM23014 treatment groups, 2) a test comparing pooled controls to the highest dose group (group 7 in mice, group 5 in rats), and 3) in mice a test comparing pooled controls to the next highest dose group (group 6). The latter pairwise test is intended to adjust for the very high mortality in the Max doses in mice. In rats only the results of the first two tests are displayed, i.e., an overall test of trend and a pairwise comparison of the pooled controls with the High dose group. Note that the number of tumors in the pooled vehicle control group is used to determine if the tumor is classified as "rare" or as "common", with the effect on interpretation as outlined below.

1.3.1.4. 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. Current FDA practice is based on the Haseman-Lin-Rahman rules. Namely, based on his extensive experience with such analyses, for pairwise tests comparing control to the highest dose group, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. For a standard chronic study in two species, i.e., rats and mice, 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. In this analysis the observed incidence in the vehicle control is used to decide if a tumor is rare or common (i.e.,

incidence ≤ 1 or >1 in the pooled controls). This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorigenicity when there actually is such a relation).

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. Since this study involved more than 50 animals per treatment group, and except for the highest dose group in mice, there were around 25 animals that survived to the end of the study, this criterion seems to have been satisfied. However, in mice, from the survival plots in Appendix 1 or the incidence tables in Sections 3.2.1.2, the maximum dose (30 mg/kg/day) seems to be associated with a lower survival than implied by this criterion.

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 following tables are copied from the Sponsor's reports and give the final weight and the final percent weight change relative to the pooled control in each study. Note that, roughly, this criterion seems to be satisfied in both mouse genders but seems to be exceeded in both rat genders.

Table 2: Relative Weight Change (compared to control)

Study 77006: Mice Group number & label	Dose Level (mg/kg/day)	Weight at study end			
		Males (g)	% from Control	Females (g)	% from control
1. & 2. Vehicle Control	0	41.11		38.78	
3. Low	0.5	41.80	+2	38.03	-2
4. Medium	1.5	41.63	+1	36.43	-6
5. Medium-high	5	41.23	0	37.38	-4
6. High	10	41.56	+1	36.14	-7
7. Max*	30	40.97	0	34.70	-11

* Group 7 males were terminated during Week 87 and females were terminated during Week 98. Group 7 animals were terminated prior to completion of the 104 week treatment period due to the severity of skin associated observations.

Table 2 (cont.): Relative Weight Change (compared to control)

Study 77005: Rats Group number & label	Dose Level (mg/kg/day)	Weight at study end			
		Males (g)	% from Control	Females (g)	% from control
1. & 2. Vehicle Control	0	532.5		333.6	
3. Low	0.1	506.5	-5	302.9	-9
4. Medium	0.2	461.7	-13	287.6	-14
5. High	0.5	398.8	-25	265.1	-21

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION**2.1. Overview**

Results from a study in SWISS CD®-1 mice and a study in Sprague-Dawley CD® rats were submitted to assess the carcinogenic potential of BIM23014.

2.2. Data Sources

SAS transport files 77006_tumor.XPT and 77005_tumor.XPT for mice and rats respectively were provided by the Sponsor to this reviewer.

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. Project 77006: A 104-Week Subcutaneous Injection Carcinogenicity Study in the Albino Mouse

MOUSE STUDY DURATION: Up to 104 Weeks.

DOSING STARTING DATE: April 3, 2003.

TERMINAL SACRIFICE: Final Necropsies on April 22, 2005.

EARLY DOSING TERMINATION: Males: Max Dose Group (30 mg/kg/day) Week 87.

Females: Max Dose Group (30 mg/kg/day) Week 98.

STUDY ENDING DATE (Final Report dated): April, 2006.

MOUSE STRAIN: SWISS CD-1® (ICR)BR Albino Mice.

ROUTE: Daily Subcutaneous Injection.

Seven treatment groups, groups 1-7, were formed for each of male and female CD-1 mice (60/gender in each of two identical vehicle control groups and 70/gender in each of five increasing BIM23014 treatment groups). Each animal was given a daily subcutaneous injection at one of six sites on the back: scapular left, scapular right, dorsal thoracic left, dorsal thoracic right, lumbar left, and lumbar right. Injections were sequentially rotated after each dose among the six injection sites. BIM23014 dose groups labeled Low, Medium, Medium-high, High, and Max dose groups (also labeled as groups 3-7, respectively) were injected with 0.5, 1.5, 5, 10, and 30 mg/kg/day of BIM23014 respectively, each in 10 mL/kg/day animal vehicle (i.e., 0.9% sodium chloride in water). Thus each predetermined dosing site was scheduled for dosing approximately 120 times during the 104 weeks of scheduled treatment. The Sponsor further states that animals were randomized to treatment stratified by body weight.

The Sponsor indicates that "The dose levels were selected according to the results of two preliminary 13-week studies, which demonstrated that firstly, 30 mg/kg/day was the maximum tolerated dose as indicated by a body weight decrease of approximately -12% in males and -29% in females and by an increased incidence of slight skin lesions at the injection sites and secondly 0.5 mg/kg/day was the NOAEL. Intermediate dose-levels were selected to cover the large dose-range between these two extremes." (page 20 of report) However, the Sponsor's report does not indicate if there was prior concurrence from the Reviewing Division or ECAC regarding the doses evaluated in the carcinogenicity study.

Animals were approximately six weeks old at first dosing. During the study, animals were housed individually. Food and water were available ad libitum, except during procedures. The Sponsor states that detailed physical examinations were made on all animals each week. Body weights were recorded weekly for the first 13 weeks, beginning approximately one week before initiation of dosing, and every 4 weeks thereafter.

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:

Simple mortality results are summarized in the following table:

Table 3: Sponsor's Summary Survival Counts

Group number & label	Dose Level (mg/kg/day)	Survival			
		Males	%	Females	%
1. Vehicle Control	0	21/60	35	23/60	38
2. Vehicle Control	0	22/60	37	21/60	35
3. Low	0.5	26/70	37	28/70	40
4. Medium	1.5	32/70	46	27/70	39
5. Medium-High	5	35/70	50	28/70	40
6. High	10	34/70	49	28/70	40
7. Max*	30	19/70	27	20/70	29

* Group 7 males were terminated during Week 87 and females were terminated during Week 98. Group 7 animals were terminated prior to completion of the 104 week treatment period due to the severity of skin associated observations.

Note there is a discrepancy between the counts in the terminal sacrifice group for females in the Max dose group, as shown in Table 7 in the FDA analysis in Section 3.2.1.2, below. Counts in the other groups agree. The significance levels of the Sponsor's tests on survival are summarized in the following Table 4.

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

Hypotheses	Males	Females
Homogeneity over all groups 1-7	<0.0001	0.0012
Trend over groups 1-7	<0.0001	0.0001
Homogeneity over controls	0.6978	0.7063
Pairwise controls vs. group 7	<0.0001	0.0007

Again, any differences between the controls are due to either randomization or severe problems with the conduct of the study. Thus, this reviewer doubts the utility of the test of homogeneity over controls. Other results seem similar to the FDA analysis summarized in Section 3.2.1.2. below. For both genders there is strong evidence of a lack of homogeneity in survival (both $p \leq 0.0012$), statistically significant evidence of a trend in mortality (both $p \leq$

0.0001), and statistically significant evidence of a difference between the pooled controls and the highest dose group, i.e. the Max group, (both $p \leq 0.0007$). From the Kaplan-Meier curves in Appendix 1, it is quite apparent that these results are due to the much lower survival in the Max dose group (Group 7).

Tumorigenicity analysis:

The Sponsor's protocol states that these are to be tested with pooled incidental and fatal tumors, incorporating an adjustment for multiplicity that seems to follow the usual Haseman-Lin-Rahman rules for a study with two species, each analyzed by gender (please see Section 1.3.1.4 above). Incidence tables are given in Appendix 3. Nominally statistically significant results are summarized in Table 5 below.

Table 5. Statistical Significances of Tests of Trend and Differences in Tumorigenicity

Males		Trend	Pairwise
I.S. dorsal thoracic left	Fibrosarcoma	0.0000	1+2 vs 7: 0.0000
	Malig. fibrous histiocytoma	0.0010	1+2 vs 7: 0.0210
I.S. dorsal thoracic right	Fibrosarcoma	0.0000	1+2 vs 7: 0.0006
	Malig. fibrous histiocytoma	0.0175	
I.S. lumbar left	Fibrosarcoma	0.0000	1+2 vs 7: 0.0000
	Malig. fibrous histiocytoma	0.0192	
I.S. lumbar right	Fibrosarcoma	0.0000	1+2 vs 7: 0.0002
I.S. scapular left	Fibrosarcoma	0.0003	1+2 vs 7: 0.0224
I.S. scapular right	Fibrosarcoma	0.0003	1+2 vs 7: 0.0113

Females		Trend	Pairwise
I.S. dorsal thoracic left	Fibrosarcoma	0.0000	1+2 vs 7: 0.0001
I.S. dorsal thoracic right	Fibrosarcoma	0.0002	1+2 vs 7: 0.0101
I.S. lumbar left	Fibrosarcoma	0.0104	
I.S. lumbar right	Fibrosarcoma	0.0003	1+2 vs 7: 0.0071
I.S. scapular left	Fibrosarcoma	0.0045	1+2 vs 7: 0.0396

From the incidences in the control groups displayed in Appendix 3, for both genders, each of the tumors above would be classified as rare tumors. At each injection site in males, and except for the scapular left site in females, the tests of trend in fibrosarcoma were all statistically significant (all trend $p \leq 0.0104$, all ≤ 0.0250 needed to apply the Haseman-Lin-Rahman rules). In male mice, at the left and right dorsal thoracic sites and the left lumbar site, tests of trend in malignant fibrous histiocytoma were also statistically significant (all three $p \leq 0.0192 \leq 0.0250$). Using Haseman's rules, the pairwise differences between the highest dose group (i.e., the Max group, group 7) were statistically significant at each site where a "p-value" is provided (since all $p \leq 0.0396 \leq 0.050$).

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 6 for male mice, Table 7 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 survived at the end of the interval.

Note that the protocol specified that when "the number of survivors in any group approaches 25 mice for a given sex prior to study termination, a decision will be made whether to advance to the terminal necropsies for the sex or group affected." Due to the severity of skin related reactions, particularly at the injection sites, in male mice the Max dose group (Group 7) males were terminated during Week 87. The Max dose group females were terminated during Week 98. The other dose groups were terminated at or after week 104. Two versions of the groupings of survival periods are provided, one where all treatment groups were assumed to be terminated at these early endpoints, the other where the Group 1-6 treatment groups were terminated at the protocol specified week 105.

Table 6. Summary of Male Mice Survival (dose/kg/day)

Period (Weeks)	Vehicle Control 1	Vehicle Control 2	Low 0.5 mg	Medium 1.5 mg	Medium-High 5 mg	High 10 mg	Maximum 30 mg
0-50	6/60 ¹ 90% ²	2/60 96.7%	3/70 95.7%	3/70 95.7%	7/70 90%	1/70 98.6%	21/70 70%
51-78	10/54 73.3%	12/58 76.7%	16/67 72.9%	10/67 81.4%	10/63 75.7%	15/69 77.1%	22/49 38.6%
79-86	5/44 65.0%	5/46 68.3%	8/51 61.4%	6/57 72.9%	8/53 64.3%	7/54 67.1%	8/27 27.1%
Terminal 87-106	39	41	43	51	45	47	19
79-91	7/44 61.7%	7/46 65.0%	13/51 54.3%	12/57 64.3%	11/53 60.0%	11/54 61.4%	
92-105	16/37 35.0%	17/39 36.7%	12/38 37.1%	13/45 45.7%	7/42 50%	9/43 48.6%	
Terminal 105-106	21	22	26	32	35	34	

¹ number deaths / number at risk

² per cent survival to end of period.

Table 7. Summary of Female Mice Survival (dose/kg/day)

Period (Weeks)	Vehicle Control 1	Vehicle Control 2	Low 0.5 mg	Medium 1.5 mg	Medium-High 5 mg	High 10 mg	Maximum 30 mg
0-50	3/60 ¹ 95% ²	2/60 96.7%	4/70 94.3%	2/70 97.1%	2/70 97.1%	2/70 97.1%	12/70 82.9%
51-78	12/57 75%	12/58 76.7%	11/66 78.6%	6/68 88.6%	12/68 80%	14/68 77.1%	13/58 64.3%
79-91	12/45 55%	13/46 55%	17/55 54.3%	17/62 64.3%	13/56 61.4%	13/54 58.6%	15/45 42.9%
92-97	6/33 45%	5/33 46.7%	8/38 42.9%	9/45 51.4%	10/43 47.1%	8/41 47.1%	14/30 22.9%
Terminal 98-106	27	28	30	36	33	33	16
92-104	10/33 38.3%	12/33 35.0%	10/38 40.0%	18/45 38.6%	15/43 40.0%	13/41 40.0%	
Terminal 105-106	23	21	28	27	28	28	

¹ number deaths / number at risk

² per cent survival to end of period.

Note again, in the tables above that the number of animals in the first labeled terminal period includes those animals that died in the first defined terminal period (weeks 87-106 for male mice and weeks 98-106 for female mice), while the later second terminal period includes those animals in groups 1-6 that were sacrificed after week 104.

For both mouse genders, the hypotheses of homogeneity in survival over all groups 1-7 is always rejected (all four Cox and Kruskal-Wallis tests, $p \leq 0.0104$), as was the hypotheses of no dose related trend (all four $p \leq 0.0004$). That is, for both genders we conclude that there is a dose related trend. In male mice the tests of departure from trend were also statistically significant (both $p \leq 0.0009$). In female mice, the tests of departure from trend were not statistically significant (both $p \geq 0.4535$). However, as shown in Appendix 1, when the Max dose group (Group 7) was deleted, in both genders, the tests of the hypotheses of homogeneity in survival over the remaining pooled controls and treatment groups 3-6 is not rejected (all Cox and Kruskal-Wallis tests, $p \geq 0.4737$). So these statistically significant differences were due primarily to the high mortality in the highest dose group (i.e., the max group, group 7). Kaplan-Meier plots comparing treatment groups in are given in Appendix 1, along with more details of the analysis.

Tumorigenicity analysis:

The statistically significant Peto mortality adjusted tests of trend in the incidence of neoplasms over the pooled controls and the five BIM23014 treatment groups, the pairwise tests of differences between pooled controls and the highest dose group (group 7, labeled Max dose), and the pairwise tests of differences between the pooled controls and the next highest dose group (group 6, labeled High dose) are all presented below. These incidence tables and statistically nonsignificant results are displayed in more detail in Appendix 3. Again, the observed

spontaneous background rate from the pooled vehicle treatment group is used to determine if the tumor is classified as a rare tumor or as a common tumor.

To adjust for the multiplicity of comparisons involved in a tumorigenicity analysis for standard rodent models, the Agency analysis followed the Haseman-Lin-Rahman rules described in Section 1.3.1.2 above. The only potentially statistically significant or close to significant tests are summarized below. Note that from the incidences in the control groups, for both genders, each of the following tumors would be classified as rare tumors. At each injection site and pooled over injection sites, the tests of trend in fibrosarcoma and, except for the left lumbar site in female mice, all pairwise comparisons between the maximum dose group and the pooled controls were statistically significant (all four trend $p \leq 0.0122 \leq 0.025$ and all relevant pairwise comparisons $p \leq 0.0247 \leq 0.05$). In male mice, at the left and right dorsal thoracic sites and the left lumbar site, tests of trend in malignant fibrous histiocytoma were all statistically significant (all four $p \leq 0.0192$). The tests of pairwise differences in malignant fibrous histiocytoma between the maximum dose group and the pooled controls in male mice was statistically significant at all pooled injection sites and at the left dorsal thoracic site ($p=0.0025$ and $p=0.0226$, respectively). In addition, deleting the maximum dose group, the test in female mice comparing all neoplasms over all injection sites at the high dose group (Group 6) to the pooled controls was statistically significant ($p = 0.0212$). Note however, that strictly speaking, the usual Haseman-Lin-Rahman rules do not apply to this comparison, and it can not necessarily be considered as adjusted for multiplicity. As noted earlier the toxicologist has the opinion that these statistically significant results are not strictly evidence of drug related carcinogenicity but are likely due to local irritation effects.

Table 8. Potentially Statistically Significant Tumorigenicity in Male Mice

	Incidence:							p-values:		
	Ctrl	Ctrl2	Low	Med	Med-High	Hi	Max	Trend	Max vs Ctrls	Hi vs Ctrls
Injection Site Dorsal Thoracic Left										
Fibrosarcoma	0	0	0	0	0	0	13	0.0000	0.0000	
Malignant fibrous histiocytoma	0	0	0	0	0	0	3	0.0010	0.0226	
Injection Site Dorsal Thoracic Right										
Fibrosarcoma	0	0	0	0	0	0	6	0.0000	0.0007	
Malignant fibrous histiocytoma	0	0	0	0	0	0	2	0.0179	0.1109	
Injection Site Lumbar, Left										
Fibrosarcoma	0	0	0	0	0	0	10	0.0000	0.0000	
Malignant fibrous histiocytoma	0	0	0	0	0	0	2	0.0192	0.1180	
Injection Site Lumbar, Right										
Fibrosarcoma	0	0	0	0	0	0	9	0.0000	0.0001	
Injection Site Scapular, Left										
Fibrosarcoma	0	0	0	0	0	1	4	0.0004	0.0272	
Injection Site Scapular, Right										
Fibrosarcoma	0	0	0	0	0	0	4	0.0004	0.0142	
Injection Site										
Any neoplasm	0	0	0	2	0	1	28	0.0000	0.0000	
Fibrosarcoma	0	0	0	0	0	1	27	0.0000	0.0000	
Malignant fibrous histiocytoma	0	0	0	0	0	0	4	0.0000	0.0025	

Table 9. Potentially Statistically Significant Tumorigenicity in Female Mice

	Incidence:							p-values:		
	Ctrl	Ctrl2	Low	Med	Med-High	Hi	Max	Trend	Max vs Ctrls	Hi vs Ctrls
Injection Site Dorsal Thoracic Left										
Fibrosarcoma	0	0	1	0	0	0	9	0.0000	0.0001	
Injection Site Dorsal Thoracic Right										
Fibrosarcoma	0	0	0	0	0	1	4	0.0003	0.0156	
Injection Site Lumbar, Left										
Fibrosarcoma	0	0	0	0	1	1	2	0.0122	0.1082	
Injection Site Lumbar, Right										
Fibrosarcoma	0	0	0	0	1	0	4	0.0004	0.0097	
Injection Site Scapular, Left										
Fibrosarcoma	0	0	1	0	0	0	3	0.0035	0.0247	
Injection Site Any neoplasm	0	0	2	1	4	4	15	0.0000	0.0000	0.0212
Fibrosarcoma	0	0	2	0	3	2	14	0.0000	0.0000	0.1505

3.2.2. Project 77005: A 104-Week Subcutaneous Injection Carcinogenicity Study of BIM 23014 in the Albino Rat

RAT STUDY DURATION: Week 104.

DOSING STARTING DATE: November 7, 2002.

TERMINAL SACRIFICE: Final necropsies: November 24, 2004.

STUDY ENDING DATE (Final Report dated): April 13, 2006.

RAT STRAIN: Sprague-Dawley CD® ..CD® (SD)BR) Rats.

ROUTE: Daily Injection.

Five treatment groups were formed for each of male and female CD-1 mice (60 animals/gender in each of two putatively identical control groups and 70 animals/gender in each of three BIM23014 treatment groups). Each animal was given a daily subcutaneous injection at one of seven sites on the back: left/right lumbar, left/ right dorsal thoracic, dorsal thoracic, and left/right scapular. As with mice, injection sites were rotated after each dose between the seven injection sites. Dose groups labeled Low, Medium, and High, (also labeled as groups 3-5, respectively) were injected with 0.1, 0.2, and 0.5 mg/kg/day BIM23014, each in 10 mL/kg/day animal vehicle (i.e., 0.9% sodium chloride in water). Thus each predetermined dosing site was scheduled for dosing approximately 104 times during the 104 weeks of scheduled treatment. The Sponsor further states that animals were randomized to treatment stratified by body weight.

The Sponsor indicates that "The dose levels were selected according to the potential human exposure, existing toxicity data and any limitations imposed by the test article" (page 21 of report). The Sponsor's report does not indicate that there was prior concurrence from the Reviewing Division or ECAC regarding the doses evaluated in the study.

During the study animals were housed individually. Water was available ad libitum. "Male and female rats were offered 5 or 4 pellets per day, respectively, of a standard commercial

laboratory diet . . . , except during designated procedures.” (page 22 of report). The Sponsor states that detailed physical examinations were made on all animals each week. Body weights were recorded weekly for the first 13 weeks, beginning approximately one week before initiation of dosing, and every 4 weeks thereafter.

3.2.2.1 Sponsor’s Results and Conclusions

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

Survival analysis:

Simple mortality results are summarized in the following table:

Table 10: Sponsor’s Summary Survival Counts

Group number & label	Dose Level (mg/kg/day)	Survival			
		Males	%	Females	%
1. Vehicle Control	0	43/60	72	34/60	57
2. Vehicle Control	0	38/60	63	25/60	42
3. Low	0.1	53/70	76	45/70	64
4. Medium	0.2	57/70	81	49/70	70
5. High	0.5	42/70	60	49/70	70

Except for a single animal, these results agree with the corresponding tables 13 and 14 reported in the FDA analysis in Section 3.2.2.2, below.

The significance levels of the Sponsor’s tests on survival are summarized in the following Table 11.

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

Hypotheses	Males	Females
Homogeneity over all groups 1-5	0.0142	0.0115
Trend over groups 1-5	0.1190	0.0080
Homogeneity over controls	0.3837	0.1532
Pairwise controls vs. group 5	0.0039	Not significant

Again, any differences between the controls are due to either randomization or severe problems with the conduct of the study. Thus, this reviewer doubts the utility of the test comparing the controls. Other results seem similar to the FDA analysis summarized in Section 3.2.2.2 below. For both genders, there was evidence of a lack of homogeneity in survival (both $p \leq 0.0142$). In female rats, there was a statistically significant evidence of a (negative) trend in mortality ($p = 0.0080$). Other comparisons were not statistically significant.

Tumorigenicity analysis:

The Sponsor's protocol states that these are to be tested with pooled incidental and fatal tumors, incorporating an adjustment for multiplicity that seems to follow the usual Haseman-Lin-Rahman rules for a study with two species by two genders (please see Section 1.3.1.4 above). The results of any potentially statistically significant trends and comparison are summarized in table 12, below:

Table 12. Statistical Significances of Tests of Trend and Differences in Tumorigenicity

Males		Trend	Pairwise
Hemolym. tissue	Malignant lymphoma	0.0152	1+2 vs 5: 0.0401
I.S. dorsal thoracic left	Fibrohistiosarcoma	0.0226	
	Fibrosarcoma	0.0064	1+2 vs 5: 0.0379
I.S. lumbar left	Fibrohistiocytic sarcoma	0.0084	1+2 vs 5: 0.0377
I.S. lumbar right	Fibrohistiocytic sarcoma	0.0063	1+2 vs 5: 0.0402
All injection sites	Combined tumors	0.0000	1+2 vs 5: 0.0000
I.S. tumor sites	Combined tumors types	0.0000	1+2 vs 5: 0.0000
Females		Trend	Pairwise
I.S. dorsal thoracic left	Fibrohistiosarcoma	0.0406	
All injection sites	Combined tumors	0.0000	1+2 vs 5: 0.0018
I.S. tumor types	Combined tumors types	0.0000	1+2 vs 5: 0.0011

Again, using these Haseman-Lin-Rahman rules, it is important to distinguish between rare and common tumors. Of these potentially statistically significant differences, due to the incidence in the pooled controls, for each gender only the groupings of "combined tumors" at any injection site would be classified as a common tumor. The remaining neoplasms cited above would be classified as rare tumors. Using these rules, in male rats both the test of trend and the test of pairwise differences between the high dose group and the pooled controls in malignant lymphoma in hemolym tissue were statistically significant (trend $p = 0.0152$ and pairwise $p = 0.0401$). Further, in male rats pooling over all injection sites, as well as the left dorsal thoracic site, and both the left and right lumbar injection sites, the tests of trend in fibrohistiocytic sarcoma were all statistically significant (all trend $p \leq 0.0226$). The corresponding pairwise comparisons at both the left and right lumbar injection sites in male rats were also statistically significant (both $p \leq 0.0402$). In the male rats, for fibrosarcoma at the left dorsal thoracic site, the test of overall trend and pairwise differences between the high dose group and the pooled controls were both statistically significant (both trend and pairwise $p = 0.0000$).

3.2.2.2 FDA Reviewer's Results

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

Survival analysis:

Again, Kaplan-Meier plots comparing survival among treatment groups in both studies are given in Appendix 1, along with more details of the analysis. The following tables (Table 13 for male rats, Table 14 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent survived to the end of the interval.

Table 13. Summary of Male Rat Survival (dose/kg/day)

Period (Weeks)	Vehicle Control 1	Vehicle Control 2	Low 0.1 mg/kg/day	Medium 0.2 mg/kg/day	High 0.5 mg/kg/day
0-50	1/60 ¹ 98.3% ²	2/60 96.7%	2/70 97.1%	0/70 100%	5/70 92.9%
51-78	3/59 93.3%	4/58 90%	3/68 92.9%	1/70 98.6%	11/65 77.1%
79-91	6/56 83.3%	0/54 90%	6/65 84.3%	6/69 90%	3/54 72.9%
92-104	7/50 71.7%	16/54 63.3%	6/59 75.7%	6/63 81.4%	9/51 60%
Terminal	43	38	53	57	42

¹ number deaths / number at risk

² per cent survival to end of period.

Table 14. Summary of Female Rat Survival (dose/kg/day)

Period (Weeks)	Vehicle Control 1	Vehicle Control 2	Low 0.1 mg/kg/day	Medium 0.2 mg/kg/day	High 0.5 mg/kg/day
0-50	0/60 ¹ 100% ²	2/60 96.7%	3/70 95.7%	1/70 98.6%	2/70 97.1%
51-78	6/60 90%	5/58 88.3%	3/67 91.4%	5/69 91.4%	4/68 91.4%
79-91	7/54 78.3%	11/53 70%	9/64 78.6%	8/64 80%	7/64 81.4%
92-104	13/47 21.7%	17/42 41.7%	11/55 62.9%	7/56 70%	8/57 70%
Terminal	34	25	44	49	49

¹ number deaths / number at risk

² per cent survival to end of period.

Although exact significance levels differ between this analysis and the Sponsor's analysis above, results are consistent. For both genders the hypotheses of homogeneity in survival across

treatment groups is always rejected (all $p \leq 0.0231$). In female rats hypotheses of no trend is also rejected (both $p \leq 0.0127$). From the Kaplan-Meier curves in Appendix 1, this is clearly a negative dose related trend. That is, the highest mortality occurs in the control groups. In male rats the tests of trend were possibly somewhat close to statistical significance (Cox $p=0.1187$ and K-W $p=0.0634$). Further details are provided in Appendix 1.

Tumorigenicity analysis:

The Peto mortality adjusted tests of trend in the incidence of neoplasms over the pooled controls and the three BIM23014 treatment groups, and the pairwise tests of differences between the pooled controls and the high dose group (group 5), and the supporting incidence tables are displayed in tables A.2.3 and A.2.4 in Appendix 3. Again, for tumor types with 10 or fewer tumor bearing animals across the treatment groups the results of an exact test (assuming fixed marginals) are provided. For tumor types with more than 10 tumor bearing animals across the treatment groups, the results from an asymptotic test are given. As noted earlier, the observed spontaneous background rate from the pooled vehicle treatment group is used to determine if the tumor is classified as a rare tumor or as a common tumor.

To adjust for the multiplicity of comparisons involved in a tumorigenicity analysis for standard rodent models, the Agency analysis followed the Haseman-Lin-Rahman rules described in Section 1.3.1.4 above. The only potentially statistically significant or close to significant tests are summarized below. Using these rules, it is important to distinguish between rare and common tumors. Of these potentially statistically significant differences, due to the incidence in the pooled controls, for each gender only the grouping of "any neoplasm" at any injection site would be classified as a common tumor. The remaining neoplasms cited below would be classified as rare tumors. Using these rules, in male rats both the test of trend and the test of pairwise differences between the high dose group and the pooled controls in malignant lymphoma in hemolymph tissue were statistically significant (trend $p = 0.0146$ and pairwise $p = 0.0389$). Further, in male rats over all injection sites, as well as the left dorsal thoracic site, and both the left and right lumbar injection sites, the tests of trend in malignant fibrous histiocytoma were all statistically significant (all trend $p \leq 0.0217$). The corresponding pairwise comparisons at both the left and right lumbar injection sites in male rats were also statistically significant (both $p \leq 0.0390$). In the male rats, for fibrosarcoma at the left dorsal thoracic site, the test of overall trend and pairwise differences between the high dose group and the pooled controls were both statistically significant (trend $p = 0.0061$ and pairwise $p = 0.0356$). Pooling all injection sites in male rats, the tests for overall trend and pairwise differences between the high dose group and the pooled controls in fibrosarcoma were both statistically significant (trend $p = 0.0010$ and pairwise $p = 0.0161$). Pooling all neoplasms over all injection sites the tests for overall trend and pairwise differences between the high dose group and the pooled controls were both statistically significant (trend $p < 0.0001$ and pairwise $p < 0.0001$). In female rats, pooling over all injection sites, the tests of trend in fibrosarcoma, malignant fibrous histiocytoma, and all neoplasms were all statistically significant (all three $p \leq 0.0023$ – but recall that the grouping of all neoplasms is not considered a rare tumor). The tests of pairwise differences in female rats in fibrosarcoma, malignant fibrous histiocytoma, and all neoplasms were all statistically significant

($p = 0.0136$, $p = 0.0210$, and $p = 0.0020$, respectively). As noted earlier the toxicologist has the opinion that these neoplasms are not strictly evidence of drug related carcinogenicity but are rather due to local irritation effects.

Table 15. Potentially Statistically Significant Tumorigenicity in Male Rats

Organ / Tumor	Incidence:					p-values:	
	Con- trol1	Con- trol2	Low	Med- ium	High	Trend	High vs Controls
HEMOLYM. TISSUE							
Malignant lymphoma	0	0	1	1	3	0.0146	0.0389
Injection Site Dorsal Thoracic Left							
Fibrosarcoma	0	0	0	0	3	0.0061	0.0356
Fibrous histiocytoma: mal	0	1	0	0	3	0.0217	0.1101
Injection Site Lumbar, Left							
Fibrous histiocytoma: mal	0	0	0	2	3	0.0084	0.0386
Injection Site Lumbar, Right							
Fibrous histiocytoma: mal	0	0	0	0	3	0.0059	0.0390
Injection Site							
Any neoplasm	1	4	2	6	16	0.0000	0.0000
Fibrosarcoma	0	1	0	0	5	0.0010	0.0161
Fibrous histiocytoma: mal	0	1	0	2	10	0.0000	0.0000

Table 16. Potentially Statistically Significant Tumorigenicity in Female Rats

Organ / Tumor	Incidence:					p-values:	
	Con- trol1	Con- trol2	Low	Med- ium	High	Trend	High vs Controls
Injection Site Dorsal Thoracic Left							
Fibrosarcoma	0	0	0	0	2	0.0567	0.1725
Fibrous histiocytoma: mal	0	0	0	0	2	0.0474	0.1423
Injection Site							
Any neoplasm	0	2	0	1	9	0.0001	0.0020
Fibrosarcoma	0	0	0	0	5	0.0008	0.0136
Fibrous histiocytoma: mal	0	0	0	0	4	0.0023	0.0210

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.

APPENDICES:**Appendix 1. Survival Analysis**

The statistical significance of the tests of differences in survival across treatment groups are given below. Since differences between the two vehicle controls should be solely due to randomization, for the tests below these two control groups are pooled to a single control group. The test for homogeneity is a test that survival is equal across treatment groups, while the test of trend is a test of dose related trend. Note that the Cox test is usually called the logrank test, while the K-W, i.e., Kruskal-Wallis test, is more commonly called the Wilcoxon test.

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

	Mice: Males		Females		Rat: Males		Females	
	Cox	K-W	Cox	K-W	Cox	K-W	Cox	K-W
Homogeneity	<0.0001	<0.0001	0.0104	0.0062	0.0226	0.0141	0.0082	0.0231
Trend over all groups	<0.0001	<0.0001	0.0004	0.0003	0.1187	0.0634	0.0080	0.0127
Departure from trend	<0.0001	0.0009	0.5507	0.4535	0.0300	0.0288	0.0824	0.1628

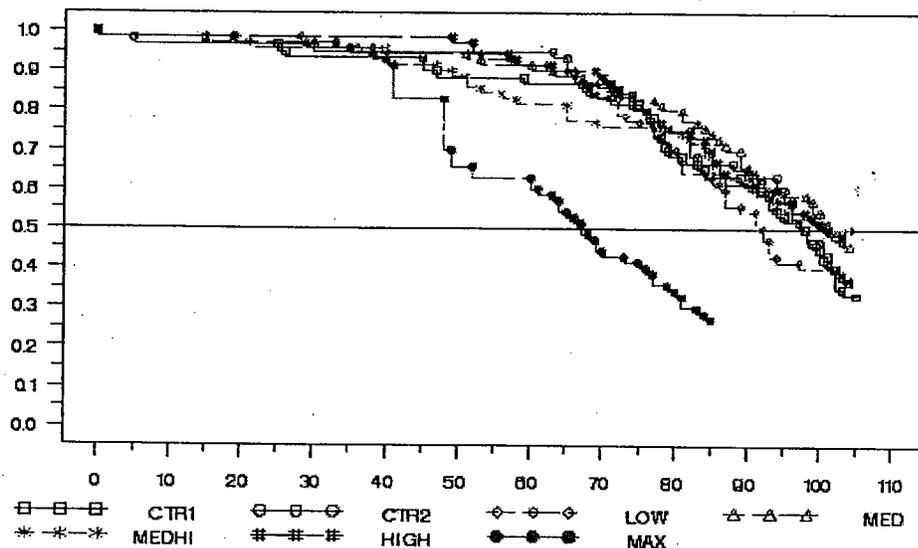
Mice Deleting Maximum Dose Group

Homogeneity	0.4737	0.6283	0.9658	0.8148
Trend over groups 1-6	0.1164	0.2210	0.7543	0.8397
Departure from trend	0.7206	0.7414	0.9301	0.6989

Over all treatment groups, for both species and genders the hypotheses of homogeneity is always rejected (all $p \leq 0.0231$). However, in both genders in mice, after deleting the maximum dose group, the hypothesis of homogeneity among the remaining six dose groups can not be rejected (all $p \geq 0.4737$). For both genders in mice and in female rats, the hypotheses of no trend over all groups are also rejected (all six $p \leq 0.0127$). The corresponding tests for male rats are arguably somewhat close to statistical significance (Cox $p=0.1187$ and K-W $p=0.0634$). In male rats and especially male mice, the tests of lack of homogeneity above and beyond that explained by trend over groups were all statistically significant (all four $p \leq 0.0300$). In female mice, the tests of such departures from trend were quite statistically nonsignificant (both $p > 0.4535$), but possibly somewhat close to significance in female rats (Cox $p=0.1187$ and K-W $p=0.0634$). From the incidence tables (tables 6, 7, 13, and 14) or the Kaplan-Meier curves below, one can see that in mice and male rats, the highest dose group has the highest mortality, with the remaining dose groups relatively closely intertwined. However, in female rats, the control groups seem to have the highest mortality, with the three BIM23014 dose groups relatively intertwined with lower mortality.

The figures below display these Kaplan-Meier estimated survival curves for the two genders in each rodent species.

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



For female mice the survival plots intertwine as depicted below:

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

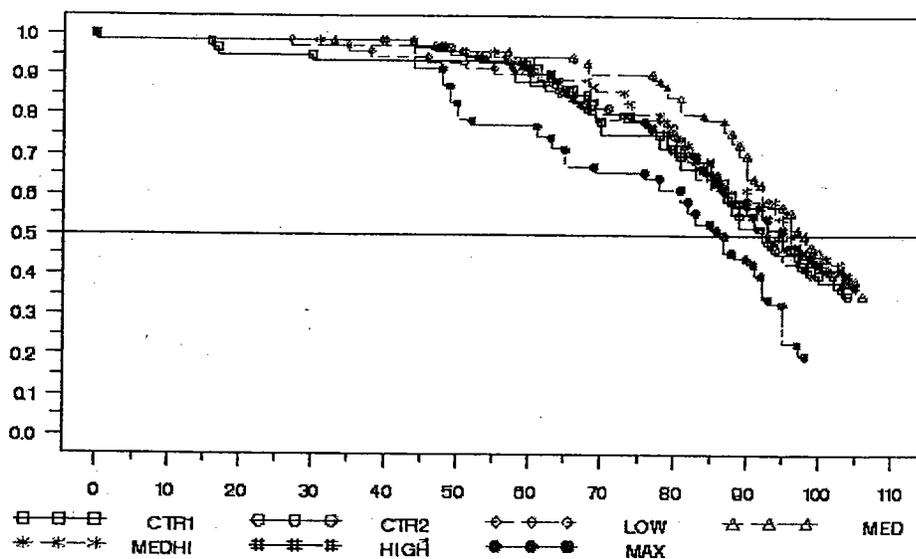


Figure A.1.3 Kaplan-Meier Survival Curves for Male Rats

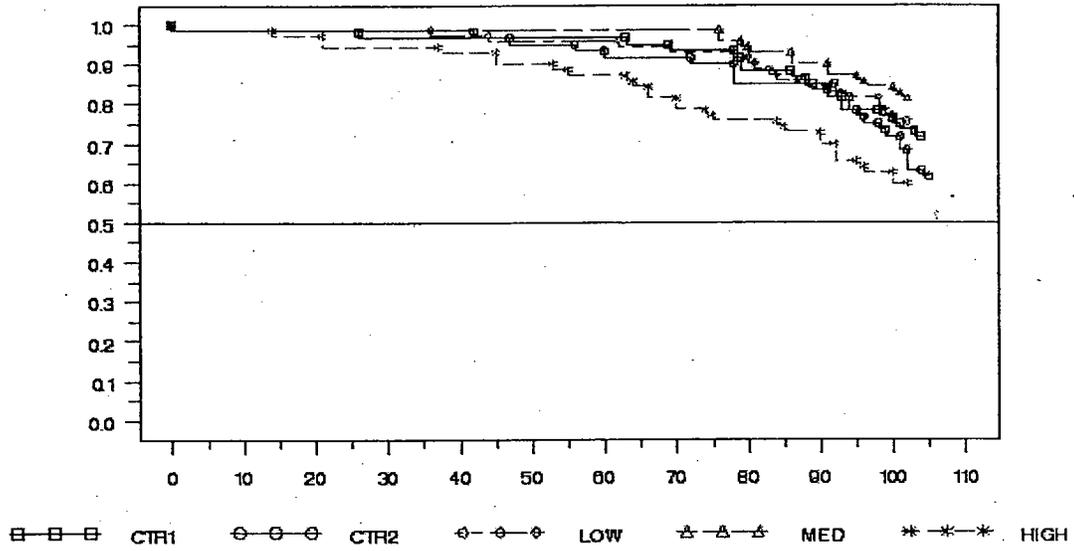
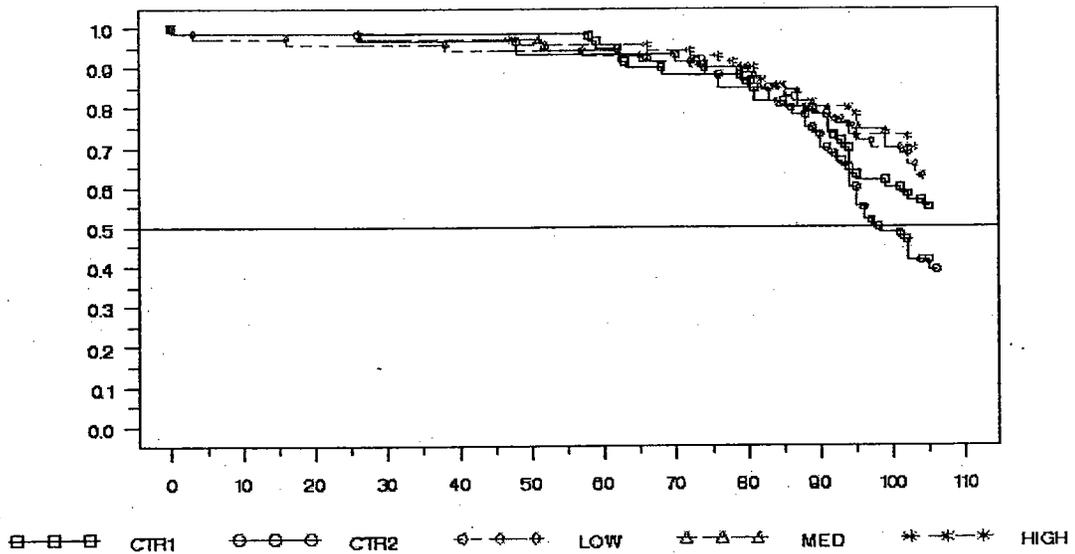


Figure A.1.4 Kaplan-Meier Survival Curves for Female Rats



Appendix 2. Bayesian Analysis of Survival

Let $S(t)$ be the survival function, i.e., with T denoting the survival function,
 $S(t) = Pr(T > t)$,
 and $f(t)$ the density of T . The instantaneous hazard function is $h(t) = f(t)/S(t)$ with cumulative hazard:

$$H(t_i) = \int_0^{t_i} h(u) du$$

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:

$$h(t | x) = h_0(t) \exp(x'\beta).$$

Frequentist analysis of this model uses asymptotics to analyze the linear predictor, ignoring the baseline hazard $h_0(t)$. A Bayesian analysis requires priors on all parameters, including the baseline hazard. Perhaps the simplest Bayesian model would postulate a within interval constant baseline hazard. For this analysis, the intervals were chosen as (0,380], (380,500], (500,580], (580,640], and (640,terminal]. This analysis assumes a within interval constant baseline hazard.

Thus we need to specify an appropriate prior for the baseline hazard. Note that the baseline hazard is essentially the hazard of the control group. An unbounded uniform prior on the baseline hazards is improper but, at least in this case, results in a proper posterior distribution, and, partly for experimental reasons, was chosen as the prior for this analysis. The priors on regression parameters were a well dispersed normal distribution (i.e., $N(0.0, 100,000)$).

In mice there were nominally seven treatment groups and in rats there were five, including two putatively identical controls. Unless there were severe structural problems with the studies, any differences between the two vehicle controls should be due solely to randomization. Thus, in this analysis the two controls were pooled, resulting in six dose groups in mice and four in rats. In the formulation above, the baseline hazard is partially confounded with the specification of treatment effects (i.e., a multiplicative constant can be moved to either the baseline hazard or the term with covariates). So there are only five degrees of freedom for testing differences among the six treatment groups in mice and three degrees of freedom for testing differences among the four treatment groups in rats.

When parameterizing each treatment group separately, using so called dummy coding, we can define, for each treatment group i , except the highest dose:

$$\delta_i = \begin{cases} 1 & \text{for the } i\text{th treatment group,} \\ 0 & \text{otherwise.} \end{cases}$$

With this parameterization each labeled effect actually represents the differential effect of the specified treatment over the effect of the highest dose group.

Three or four possibly relevant models for treatment effect could be expressed as follows:

(1) Parameterization of a differential effect over the last treatment, labeled k (with k=6 treatments in mice, k=4 in rats, including pooled controls),

$$x_i^t \beta = \beta_0 + \beta_1 * \delta_1 + \beta_2 * \delta_2 + \dots + \beta_{k-1} * \delta_{k-1}.$$

(2) Parameterization of a linear effect of measures dose over treatment groups with vehicle,

$$x_i^t \beta = \beta_0 + \beta_1 * \text{dose}.$$

(3) Parameterization of no differences in survival across treatment groups with vehicle, (i.e., constant dose effect) $x_i^t \beta = \beta_0$.

(4) Parameterization with only a difference between the maximum dose group versus controls, and versus the pooled remaining dose groups, with $x_i^t \beta = \beta_0 + \beta_1 * \delta_1 + \beta_5 * \delta_5$ (mice only).

Note again, that for each of these models $\exp(\beta_0)$ is confounded with the baseline hazard $h_0(t)$ and is not estimated. The early sacrifice in mice will tend to reduce the number of events in the maximum dose group, and thus we will tend to underestimate the treatment effect on survival in this dose group. In model (4) above, β_1 measures the differences between the maximum dose and the pooled controls, while the β_5 measures the differences between the maximum dose and the pooled remaining BIM23014 dose groups. The program used for this analysis was the experimental SAS® procedure, PROC BPHREG. Because this is a new procedure and is still considered to be experimental, this analysis, at best, can only be considered to be supporting.

One approach for model selection in Bayesian models is to use the Deviance Information Criterion (DIC). Effectively, for $D(\theta)$ denoting the usual deviance, $DIC \approx E(D(\theta)) + 1/2 (\text{Var}(D(\theta)))$. For good models we would want the deviance and the variance to be as small as possible. Thus, for a given data set the model with the smallest DIC would be preferred. The estimated DICs are given below:

Deviance Information Criterion for Mice	Males	Females
Model with heterogeneity over all dose groups	4094.15	4245.46
Model with pooled controls, maximum dose, & with other doses equal	4090.97	4239.85
Model with linear trend in dose	4091.56	4232.92
Model with constant dose effect (i.e., equivalently, no dose effect)	4122.61	4248.69

Deviance Information Criterion for Rats	Males	Females
Model with heterogeneity over all dose groups	1526.21	1926.82
Model with linear trend	1525.38	1932.61
Model with constant dose effect (i.e., equivalently, no dose effect)	1526.77	1936.98

Among the models assessed, for male mice, the best fitting model has an effect for the maximum dose group, the pooled controls, with the remaining dose groups pooled. However,

the model with a linear effect in dose is almost as good, while for female mice it is the best fitting model. In male rats, the DIC indicates that the model with linear trend in dose is slightly better than the others, while for female rats the model with a different effect for each treatment group is the best of the three models.

Table A.2.1 for mice and table A.2.2 for rats, below, summarize the estimated posterior distributions of the treatment group parameters. Note that for each gender the approximate 95% credible intervals for both the parameters corresponding to the linear effect of dose and the parameters corresponding to the differences between BIM23014 doses and the vehicle control include zero well within the intervals. This can be interpreted as being consistent with the hypothesis that these parameters are also zero, confirmation of the conclusions based on the DIC.

Table A.2.1 Posterior Summaries of Treatment Parameters in the Mice Study

Parameter	Mean	Standard Deviation	Quantiles			HPD 95% Credible Interval	
			25%	50%	75%		
Male Mice: Dose groups differ							
Dose Grp=0	-1.0940	0.1898	-1.2223	-1.0946	-0.9663	-1.4496	-0.7052
Dose Grp=1	-1.0592	0.2146	-1.2037	-1.0570	-0.9147	-1.4859	-0.6434
Dose Grp=2	-1.3439	0.2231	-1.4925	-1.3438	-1.1925	-1.7785	-0.9083
Dose Grp=3	-1.3743	0.2281	-1.5269	-1.3727	-1.2200	-1.8233	-0.9317
Dose Grp=4	-1.4080	0.2273	-1.5598	-1.4059	-1.2547	-1.8540	-0.9594
Male Mice: Linear trend in dose							
Linear Dose	0.0365	0.00617	0.0324	0.0366	0.0407	0.0243	0.0486
Male Mice: Three groups: controls, maximum dose, other							
Dose Grp=0	-1.0880	0.1883	-1.2153	-1.0898	-0.9610	-1.4515	-0.7148
Dose Grp=1-4	-1.2788	0.1709	-1.3957	-1.2802	-1.1652	-1.6169	-0.9471
Female Mice: Dose groups differ							
Dose Grp=0	-0.6202	0.1767	-0.7399	-0.6199	-0.5012	-0.9554	-0.2621
Dose Grp=1	-0.6840	0.2054	-0.8203	-0.6813	-0.5473	-1.0897	-0.2817
Dose Grp=2	-0.7771	0.2050	-0.9141	-0.7761	-0.6385	-1.1714	-0.3711
Dose Grp=3	-0.7259	0.2053	-0.8630	-0.7242	-0.5874	-1.1293	-0.3244
Dose Grp=4	-0.6515	0.2017	-0.7879	-0.6495	-0.5164	-1.0392	-0.2479
Female Mice: Linear trend in dose							
Linear Dose	0.0228	0.00542	0.0191	0.0228	0.0265	0.0123	0.0334
Female Mice: Three groups: controls, maximum dose, other							
Dose Grp=0	-0.6214	0.1776	-0.7413	-0.6215	-0.5010	-0.9638	-0.2704
Dose Grp=1-4	-0.7027	0.1545	-0.8079	-0.7047	-0.5982	-1.0112	-0.4092

Note that in each model, the 95% credible intervals for the parameters never contain 0. The credible intervals for the models for trend are positive, indicating an increase in mortality due to dose. For the models with simple treatment effects, whether a different effect for each dose or the model with pooled controls and, except for the maximum dose group, pooled BIM23014 treatment groups are negative, bounded away from 0. In each model, this indicates that the maximum dose group has significantly the highest mortality.

Table A.2.2 Posterior Summaries of Treatment Parameters in the Rat Study

Parameter	Mean	Standard Deviation	Quantiles			HPD 95%	
			25%	50%	75%	Credible Interval	
Male Rats: Dose groups differ							
Dose Grp=0	-0.3238	0.2482	-0.4924	-0.3276	-0.1568	-0.8107	0.1624
Dose Grp=1	-0.6726	0.3121	-0.8831	-0.6697	-0.4618	-1.2893	-0.0641
Dose Grp=2	-1.0052	0.3397	-1.2322	-1.0015	-0.7730	-1.6693	-0.3467
Male Rats: Linear trend in dose							
Linear Dose	0.0758	0.0536	0.0400	0.0761	0.1125	-0.0323	0.1772
Female Rats: Dose groups differ							
Dose Grp=0	0.7096	0.2539	0.5354	0.7047	0.8760	0.2292	1.2147
Dose Grp=1	0.2678	0.2961	0.0685	0.2658	0.4651	-0.3321	0.8303
Dose Grp=2	0.00985	0.3112	-0.1978	0.00896	0.2185	-0.6024	0.6188
Female Rats: Linear trend in dose							
Linear Dose	-0.1529	0.0539	-0.1888	-0.1515	-0.1156	-0.2582	-0.0488

When 0 is in the credible interval associated with a parameter, it can be interpreted as suggesting we can not preclude that the parameter is not 0. That is, the parameter could be 0. For male rats the credible interval for the trend parameter as well as the difference between pooled controls and the maximum dose group includes zero. Using this measure in this model, we could conclude that these parameters could be 0. Interestingly, the low and medium dose groups do seem to have lower mortality than the high dose group. In female rats the credible interval does not include zero, but is negative, indicating a higher mortality in the controls than in the high dose group.

**APPEARS THIS WAY
ON ORIGINAL**

Appendix 3. FDA Tumorigenicity Analysis

Tables A.3.1 and A.3.2 below display the number of neoplasms in each organ and tumor combination in male and female mice, respectively, while tables A.2.3 and A.2.4 present similar results in male and female rats. These values are taken from the SAS datasets provided by the Sponsor. For each dose group, the tumor incidence is the number of animals where histopathological analysis detected a tumor. The Sponsor indicates that for all tumors specified in the protocol, all animals in each treatment group were microscopically examined. In mice, three p-values of tests of hypotheses for each tumor by gender combination are presented. The column labeled "Trend" provides the observed p-value of the tests of trend over the pooled vehicle controls, and the low, medium, medium-high, high and max dose groups. The columns labeled "Max vs Ctrls" and "High vs Ctrls" provides the significance levels of the tests comparing the maximum dose group (group 7) and the high dose group (group 6) to the pooled control group. For 10 or fewer tumor bearing animals in the comparison, the reported significance levels come from exact tests (i.e., assuming that the marginal totals for the number of animals with and without the neoplasm are fixed). For more than 10 tumor bearing animals the tests large sample, asymptotic tests are used.

The Haseman-Lin-Rahman rules summarized below are designed to adjust for the multiplicity of tests over the organ by tumor combinations and determine if the observed p-value is statistically significant. That is, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level to the appropriate bound below:

Haseman - Lin - Rahman Bounds: Comparison	Rare Tumor (Incidence \leq 1%)	Common Tumor (Incidence $>$ 1%)
Trend (over 3 or more groups)	0.025	0.005
Pairwise	0.05	0.01

So, for example, for a rare tumor (with incidence in the pooled control groups \leq 1%, i.e., 0 or 1 tumor), a pairwise test between the high dose group and control would be considered statistically significant if the computed significance level was at or less than 0.05.

Table A.3.1. Tumorigenicity in Male Mice

	Incidence:							p-values:		
	Ctrl	Ctr2	Low	Med	Med-High	Hi	Max	Trend	Max vs Ctrls	Hi vs Ctrls
ADRENAL										
Adenoma: cortical	2	0	1	0	0	0	0	0.9809	1.0000	1.0000
Adenoma: subcapsular	0	4	2	4	2	1	2	0.3150	0.2869	0.8999
Benign pheochromocytoma	0	0	0	1	0	0	1	0.0649	0.1771	
Carcinoma: cortical	0	0	0	0	0	1	0	0.2238		0.3629
BONE MARROW										
Mast cell tumor (malignant)	0	0	0	0	0	1	0	0.2242		0.3651
CECUM										
Adenocarcinoma	2	0	0	0	0	0	0	1.0000	1.0000	1.0000
EPIDIDYMISS										
Carcinoma: interstitial cell	0	0	0	0	1	1	0	0.1968		0.3701
Hemangiosarcoma	0	0	0	0	1	0	0	0.5000		
Sarcoma (not otherwise speci	0	0	0	2	0	0	0	0.6548		
HARDERIAN GLAND										
Adenoma	5	3	5	5	1	7	0	0.8378	1.0000	0.2138
I.S. DORS.TH0. LT										
Fibrosarcoma	0	0	0	0	0	0	13	0.0000	0.0000	
Malignant fibrous histiocyto	0	0	0	0	0	0	3	0.0010	0.0226	
I.S. DORS.TH0. RT										
Fibrosarcoma	0	0	0	0	0	0	6	0.0000	0.0007	
Malignant fibrous histiocyto	0	0	0	0	0	0	2	0.0179	0.1109	
I.S. LUMBAR, LEFT										
Fibrosarcoma	0	0	0	0	0	0	10	0.0000	0.0000	
Malignant fibrous histiocyto	0	0	0	0	0	0	2	0.0192	0.1180	
I.S. LUMBAR, RIGHT										
Fibrosarcoma	0	0	0	0	0	0	9	0.0000	0.0001	
Malignant fibrous histiocyto	0	0	0	0	0	0	1	0.1875	0.5000	
I.S. SCAPULAR, LEF										
Fibrosarcoma	0	0	0	0	0	1	4	0.0004	0.0272	0.3651
Lipoma	0	0	0	1	0	0	0	0.5960		
Osteosarcoma	0	0	0	1	0	0	0	0.5623		
I.S. SCAPULAR, RIG										
Fibrosarcoma	0	0	0	0	0	0	4	0.0004	0.0142	
Malignant Fibrous Histiocyto	0	0	0	0	0	0	1	0.2258	0.5000	
Injection Site										
Any neoplasm	0	0	0	2	0	1	28	0.0000	0.0000	0.3701
Fibrosarcoma	0	0	0	0	0	1	27	0.0000	0.0000	0.3701
Malignant fibrous histiocyto	0	0	0	0	0	0	4	0.0000	0.0025	
JEJUNUM										
Adenocarcinoma	0	1	2	0	0	2	0	0.5354	1.0000	0.3240
KIDNEY										
Adenoma: tubular cell	0	0	0	0	0	0	1	0.2258	0.5000	
Carcinoma: tubular cell	0	1	0	0	0	0	0	1.0000	1.0000	1.0000
LIVER										
Adenoma: hepatocellular	23	18	10	17	4	15	4	0.9963	0.9995	0.9765
Carcinoma: hepatocellular	2	3	4	2	9	1	2	0.4820	0.4211	0.9488
Hemangiosarcoma	3	2	4	3	3	3	0	0.8619	1.0000	0.6788
LUNG										
Adenoma: alveolar/bronchiola	5	10	17	12	6	13	9	0.3519	0.1795	0.1198
Carcinoma: alveolar/bronchio	4	6	3	7	7	11	2	0.3461	0.4874	0.0774
PANCREAS										
Adenoma: islet cell	0	1	0	0	0	0	0	1.0000	1.0000	1.0000
PARATHYROID GLAND										
Adenoma	0	0	0	1	0	0	0	0.5698		

Table A.3.1. (cont.) Tumorigenicity in Male Mice

	Incidence:							p-values:		
	Ctrl	Ctrl2	Low	Med	Med- High	Hi	Max	Trend	Max vs Ctrls	Hi vs Ctrls
PITUITARY										
Adenoma: pars distalis	0	1	0	0	0	0	1	0.1128	0.3070	1.0000
Adenoma: pars intermedia	0	0	1	0	0	0	0	0.7160		
PREPUTIAL GLAND										
Hemangioma	0	0	0	1	0	0	0	0.5616		
PROSTATE										
Adenocarcinoma	0	0	0	0	1	1	0	0.1955		0.3651
SKIN MISCELLANEOUS										
Mast cell tumor	0	0	0	0	0	0	1	0.2000	0.7500	
SPLEEN										
Hemangioma	1	0	1	2	0	0	0	0.9406	1.0000	1.0000
Hemangiosarcoma	5	2	6	2	4	3	0	0.8970	1.0000	0.7961
Systemic										
Hemangioma	3	2	2	2	0	0	0	0.9983	1.0000	1.0000
Hemangioma/-sarcoma	11	6	13	7	8	6	0	0.9952	0.9898	0.8905
Hemangiosarcoma	8	4	11	5	8	6	0	0.9640	0.9489	0.7012
TESTIS										
Adenoma: interstitial cell	1	1	0	0	1	0	0	0.8114	1.0000	1.0000
THYROID										
Adenoma: follicular cell	1	0	0	0	0	0	0	1.0000	1.0000	1.0000
URINARY BLADDER										
Submucosal mesenchymal tumor	2	0	1	0	0	1	1	0.1735	0.4428	0.7536

Table A.3.2. Tumorigenicity in Female Mice

	Incidence:							p-values:		
	Ctrl	Ctrl2	Low	Med	Med- High	Hi	Max	Trend	Max vs Ctrls	Hi vs Ctrls
ADRENAL										
Adenoma: cortical	0	0	0	0	0	0	1	0.1648	0.3902	
Adenoma: subcapsular	3	1	2	2	1	0	0	0.9849	1.0000	1.0000
Benign pheochromocytoma	0	0	1	3	0	3	0	0.6166		0.0495
Carcinoma: cortical	0	0	1	0	0	0	0	0.7282		
Malignant pheochromocytoma	2	1	1	1	1	2	0	0.7501	1.0000	0.6246
BONE MARROW										
Mast cell tumor (malignant)	2	0	0	0	0	0	0	1.0000	1.0000	1.0000
BONE-FEMUR										
Chondroma	0	0	0	1	0	0	0	0.6667		
Osteosarcoma	0	0	0	1	0	0	0	0.6129		
BONE-STERNUM										
Hemangiosarcoma	0	1	0	0	0	0	0	1.0000	1.0000	1.0000
ESOPHAGUS										
Carcinoma: squamous cell	0	0	0	0	0	0	1	0.1238	0.3305	
FAT										
Liposarcoma	0	0	0	0	0	0	1	0.2000	0.2500	
HARDERIAN GLAND										
Adenoma	2	1	1	1	6	0	0	0.8383	1.0000	1.0000
HEMOLYM. TISSUE										
Histiocytic sarcoma	2	5	6	4	6	10	5	0.0898	0.0610	0.0332
I.S. DORS. THO. LT										
Fibrosarcoma	0	0	1	0	0	0	9	0.0000	0.0001	
Mast cell tumor	0	0	0	1	0	0	0	0.6667		

Table A.3.2. (cont.) Tumorigenicity in Male Mice

	Incidence:							p-values:		
	Ctrl1	Ctrl2	Low	Med	Med- High	Hi	Max	Trend	Max vs Ctrls	Hi vs Ctrls
I.S. DORS.THO. RT										
Fibrosarcoma	0	0	0	0	0	1	4	0.0003	0.0156	0.3922
Mast cell tumor	0	0	0	0	1	0	0	0.5000		
I.S. LUMBAR, LEFT										
Fibroma	0	0	0	0	1	0	0	0.4031		
Fibrosarcoma	0	0	0	0	1	1	2	0.0122	0.1082	0.3837
Osteosarcoma	0	0	0	0	0	1	0	0.2718		0.3750
I.S. LUMBAR, RIGHT										
Fibrosarcoma	0	0	0	0	1	0	4	0.0004	0.0097	
I.S. SCAPULAR, LEF										
Fibroma	0	0	0	0	0	1	0	0.2500		0.3837
Fibrosarcoma	0	0	1	0	0	0	3	0.0035	0.0247	
Hemangioma	0	0	0	0	0	0	1	0.2000	0.4800	
Malignant fibrous histiocyto	0	0	0	0	0	0	1	0.1503	0.3730	
I.S. SCAPULAR, RIG										
Fibrosarcoma	0	0	0	0	1	0	1	0.0590	0.2319	
Injection Site										
Any neoplasm	0	0	2	1	4	4	15	0.0000	0.0000	0.0212
Fibroma	0	0	0	0	1	1	0	0.2364		0.3837
Fibrosarcoma	0	0	2	0	3	2	14	0.0000	0.0000	0.1505
Malignant fibrous histiocyto	0	0	0	0	0	0	1	0.1503	0.3730	
JEJUNUM										
Adenocarcinoma	1	0	0	0	0	1	0	0.4796	1.0000	0.6204
LIVER										
Adenoma: hepatocellular	2	2	3	4	1	2	0	0.9383	1.0000	0.7215
Carcinoma: hepatocellular	1	0	0	0	1	0	1	0.1521	0.4126	1.0000
Hemangiosarcoma	0	4	1	3	1	0	0	0.9817	1.0000	1.0000
LUNG										
Adenoma: alveolar/bronchiola	10	7	12	5	6	8	3	0.9667	0.9667	0.7217
Carcinoma: alveolar/bronchio	4	5	5	4	3	4	1	0.9097	0.9610	0.7023
Sarcoma: metastasis	0	0	0	0	0	1	0	0.2747		0.3590
LYMPH NODE										
Hemangiosarcoma	1	0	0	0	0	0	0	1.0000	1.0000	1.0000
MAMMARY GLAND										
Adenocarcinoma	4	1	4	2	4	0	0	0.9650	1.0000	1.0000
Adenoma	1	2	0	0	0	0	0	1.0000	1.0000	1.0000
OVARY										
Adenoma: tubulostromal	1	0	0	0	0	0	0	1.0000	1.0000	1.0000
Benign granulosa-theca cell	0	0	2	0	0	1	0	0.5181		0.3837
Benign sertoli cell tumor	0	0	0	0	1	0	0	0.4066		
Cystadenocarcinoma	0	0	1	0	0	0	0	0.7296		
Cystadenoma	1	0	1	0	1	1	0	0.5415	1.0000	0.6229
Hemangioma	0	0	0	0	0	1	0	0.2500		0.3837
Malignant granulosa-theca ce	1	0	0	0	1	2	0	0.3961	1.0000	0.3239
PANCREAS										
Adenoma: islet cell	0	1	0	1	0	0	0	0.8608	1.0000	1.0000
Mesothelioma(B)	0	0	1	0	0	0	0	0.7582		
PITUITARY										
Adenoma: pars distalis	1	2	3	2	1	0	0	0.9828	1.0000	1.0000
Adenoma: pars intermedia	0	0	0	1	0	0	0	0.6009		
Hemangiosarcoma	0	1	1	0	3	1	2	0.1006	0.3605	0.5994
STOMACH										
Adenoma	0	0	1	0	0	0	0	0.7296		
Mesothelioma(B)	0	0	1	0	0	0	0	0.7582		
Osteosarcoma	0	0	0	0	0	1	0	0.2500		0.3837

Table A.3.2. (cont.) Tumorigenicity in Female Mice

	Incidence:						Max	Trend	p-values:		
	Ctrl	Ctrl2	Low	Med	Med-High	Hi			Max vs Ctrls	Hi vs Ctrls	
SUBCUTANEOUS TISSUE											
Hemangiosarcoma	0	1	0	0	0	0	0	1.0000	1.0000	1.0000	
Osteosarcoma	1	0	0	0	1	0	0	0.5769	1.0000	1.0000	
Systemic											
Hemangioma	1	0	1	0	0	2	1	0.1401	0.6006	0.3268	
Hemangioma/-sarcoma	3	9	3	4	6	6	2	0.7490	0.8931	0.6796	
Hemangiosarcoma	2	9	2	4	6	4	2	0.7554	0.8705	0.8317	
THYROID											
Adenoma: follicular cell	1	0	0	2	1	0	0	0.7763	1.0000	1.0000	
URINARY BLADDER											
Carcinoma: transitional cell	0	0	0	1	0	0	0	0.5816			
Leiomyosarcoma	0	0	1	0	0	0	0	0.7296			
Submucosal mesenchymal tumor	0	0	1	0	1	0	0	0.6159			
UTERUS											
Adenocarcinoma: endometrial	0	3	7	4	5	1	4	0.1933	0.1025	0.8621	
Adenoma: endometrial	0	0	0	1	0	0	0	0.5816			
Benign granular cell tumor	1	0	1	1	2	0	1	0.4666	0.6829	1.0000	
Carcinoma: squamous cell	0	1	0	0	0	0	0	1.0000	1.0000	1.0000	
Hemangioma	0	0	0	0	0	1	0	0.2500		0.3837	
Hemangiosarcoma	1	2	0	0	2	3	0	0.5688	1.0000	0.4140	
Leiomyoma	6	3	4	7	4	2	3	0.8270	0.7457	0.9178	
Leiomyosarcoma	3	0	2	2	3	1	1	0.7077	0.8675	0.8525	
Polyp: endometrial stromal	6	6	6	9	9	7	5	0.5336	0.4579	0.5509	
Sarcoma: endometrial stromal	1	1	0	1	3	1	1	0.3518	0.6418	0.7531	
VAGINA											
Polyp	1	0	0	0	0	0	0	1.0000	1.0000	1.0000	

Table A.3.3. Tumorigenicity in Male Rats

	Incidence:					Trend	High vs. Controls
	Con-troll	Con-troll2	Low	Med-ium	High		
ADRENAL							
Benign pheochromocytoma	3	4	0	2	2	0.7711	0.8752
Malignant pheochromocytom	0	1	0	0	0	1.0000	1.0000
Malignant pheochromocytom	0	1	0	1	0	0.7227	1.0000
AORTA							
Leiomyosarcoma	0	1	0	0	0	1.0000	1.0000
BRAIN							
Malignant astrocytoma	0	0	1	1	1	0.1670	0.3308
Malignant meningioma	1	0	0	0	0	1.0000	1.0000
Malignant mixed glioma	0	2	0	1	0	0.8617	1.0000
HEMOLYM. TISSUE							
Histiocytic sarcoma	1	2	1	2	3	0.1665	0.3290
Malignant lymphoma	0	0	1	1	3	0.0146	0.0389

Table A.3.3. (cont.) Tumorigenicity in Male Rats

	Incidence:					p-values:	
	Con- troll1	Con- troll2	Low	Med- ium	High	Trend	High vs. Controls
I.S. DORS.THO. LT							
Fibroma	0	0	1	1	0	0.4909	
Fibrosarcoma	0	0	0	0	3	0.0061	0.0356
Fibrous histiocytoma: mal	0	1	0	0	3	0.0217	0.1101
Lipoma	0	1	1	0	0	0.8681	1.0000
Papilloma: epidermal	0	0	0	1	0	0.4234	
I.S. DORS.THO. RT							
Fibroma	0	0	0	1	0	0.4103	
I.S. DORSAL THORAC							
Fibroma	0	1	0	0	0	1.0000	1.0000
Fibrous histiocytoma: mal	0	0	0	0	1	0.1837	0.3250
Keratoacanthoma	0	0	0	1	0	0.4234	
I.S. LUMBAR, LEFT							
Fibrosarcoma	0	1	0	0	2	0.0838	0.2453
Fibrous histiocytoma: mal	0	0	0	2	3	0.0084	0.0386
Hemangiosarcoma	0	0	0	0	1	0.1780	0.3307
Myxoma	0	0	0	0	1	0.1889	0.3372
I.S. LUMBAR, RIGHT							
Adenoma: sebaceous gland	1	0	0	0	0	1.0000	1.0000
Fibrosarcoma	0	0	0	0	1	0.1760	0.3176
Fibrous histiocytoma: mal	0	0	0	0	3	0.0059	0.0390
Injection Site							
Any neoplasm	1	4	2	6	16	0.0000	0.0000
Fibroma	0	1	1	2	0	0.6583	1.0000
Fibrosarcoma	0	1	0	0	5	0.0010	0.0161
Fibrous histiocytoma: mal	0	1	0	2	10	0.0000	0.0000
KIDNEY							
Adenoma: tubular cell	0	0	1	0	0	0.6577	
L.NODE MESENTERIC							
Hemangioma	0	0	1	1	0	0.4298	
LIVER							
Adenoma: hepatocellular	0	1	2	0	0	0.8785	1.0000
Carcinoma: hepatocellular	1	2	0	2	0	0.8740	1.0000
LUNG							
Adenoma: alveolar/bronchi	1	0	1	0	0	0.8838	1.0000
PANCREAS							
Adenoma: islet cell	2	3	2	2	0	0.9634	1.0000
Carcinoma: islet cell	1	2	1	2	1	0.5864	0.7796
PITUITARY							
Adenoma: pars distalis	35	37	34	32	14	1.0000	1.0000
Carcinoma: pars distalis	0	1	1	0	0	0.8702	1.0000
PROSTATE							
Adenocarcinoma	0	0	0	1	0	0.4234	
SKIN MISCELLANEOUS							
Carcinoma: basal cell	0	0	0	1	0	0.5043	
Papilloma: squamous cell	1	0	0	0	1	0.4229	0.7368
SPINAL CORD CERVIC							
Malignant astrocytoma	0	0	2	0	0	0.7147	
Malignant mixed glioma	1	0	0	0	0	1.0000	1.0000
SPLEEN							
Hemangioma	0	1	0	0	0	1.0000	1.0000
Hemangiosarcoma	0	1	0	0	0	1.0000	1.0000

Table A.3.3. (cont.) Tumorigenicity in Male Rats

	Incidence:					p-values:	
	Con- trol1	Con- trol2	Low	Med- ium	High	Trend	High vs. Controls
Systemic							
Hemangioma	0	1	1	1	0	0.6527	1.0000
Hemangioma/-sarcoma	0	2	2	1	1	0.5308	0.7090
Hemangiosarcoma	0	1	1	0	1	0.4030	0.5577
TESTIS							
Adenoma: interstitial cel	1	2	1	5	1	0.4910	0.8144
THYROID							
Adenoma: C-cell	2	5	8	1	1	0.9434	0.9458
Adenoma: follicular cell	2	0	1	1	0	0.8221	1.0000
Carcinoma: C-cell	0	0	0	0	1	0.1757	0.3391
Carcinoma: follicular cel	1	1	0	0	0	1.0000	1.0000
TONGUE							
Hemangiosarcoma	0	0	1	0	0	0.6347	

Table A.3.4. Tumorigenicity in Female Rats

	Incidence:					p-values:	
	Con- trol1	Con- trol2	Low	Med- ium	High	Trend	High vs. Controls
ADRENAL							
Adenoma: cortical	1	1	0	0	0	1.0000	1.0000
Benign pheochromocytoma	1	1	1	0	1	0.6465	0.7446
Malignant pheochromocytom	0	0	1	1	0	0.6515	
BRAIN							
Malignant astrocytoma	0	2	0	0	0	1.0000	1.0000
Malignant meningioma	0	0	0	0	1	0.2893	0.5147
HEMOLYM. TISSUE							
Histiocytic sarcoma	3	4	2	0	0	0.9999	1.0000
Malignant lymphoma	0	1	0	1	0	0.7543	1.0000
I.S. DORS.TH0. LT							
Fibrosarcoma	0	0	0	0	2	0.0567	0.1725
Fibrous histiocyto: mal	0	0	0	0	2	0.0474	0.1423
I.S. DORS.TH0. RT							
Fibroma	0	1	0	0	0	1.0000	1.0000
Fibrosarcoma	0	0	0	0	1	0.2563	0.4659
Squamous cell carcinoma	0	1	0	0	0	1.0000	1.0000
I.S. DORSAL THORAC							
Fibroma	0	0	0	1	0	0.5372	
Fibrous histiocyto: mal	0	0	0	0	1	0.2215	0.3879
I.S. LUMBAR, LEFT							
Fibrosarcoma	0	0	0	0	1	0.2500	0.4369
Fibrous histiocyto: mal	0	0	0	0	1	0.2235	0.3882
I.S. SCAPULAR, LEF							
Fibrosarcoma	0	0	0	0	1	0.2248	0.3867
Fibrous histiocyto: mal	0	0	0	0	1	0.1923	0.3333
I.S. SCAPULAR, RIG							
Fibrous histiocyto: mal	0	0	0	0	1	0.2213	0.3803
Injection Site							
Any neoplasm	0	2	0	1	9	0.0001	0.0020
Fibroma	0	1	0	1	0	0.7549	1.0000
Fibrosarcoma	0	0	0	0	5	0.0008	0.0136
Fibrous histiocyto: mal	0	0	0	0	4	0.0023	0.0210

Table A.3.4. (cont.) Tumorigenicity in Female Rats

	Incidence:					p-values:	
	Con- trol1	Con- trol2	Low	Med- ium	High	Trend	High vs. Controls
JEJUNUM							
Adenocarcinoma	0	1	0	0	0	1.0000	1.0000
Fibroma	0	0	0	1	0	0.5372	
Hemangiosarcoma	0	0	1	0	0	0.5593	
KIDNEY							
Renal mesenchymal tumor	0	1	0	0	0	1.0000	1.0000
LIVER							
Adenoma: hepatocellular	1	1	1	4	1	0.7362	0.8911
Carcinoma: hepatocellular	0	1	0	0	0	1.0000	1.0000
Cholangioma	0	1	0	0	0	1.0000	1.0000
MAMMARY GLAND							
Adenocarcinoma	10	16	13	5	2	1.0000	0.9999
Adenoma	10	10	15	5	2	0.9999	0.9997
Carcinosarcoma	1	0	1	2	0	0.7972	1.0000
Fibroadenoma	13	11	13	12	10	0.9536	0.9303
Fibroma	1	0	2	0	0	0.8892	1.0000
OVARY							
Adenoma: sertoliform tubu	1	0	0	0	0	1.0000	1.0000
Adenoma: tubulostromal	0	0	0	1	0	0.5372	
Carcinoma: sertoliform	0	0	0	0	1	0.2893	0.5147
Cystadenoma	2	0	0	0	0	1.0000	1.0000
Malignant granulosa-theca	0	1	1	1	0	0.8549	1.0000
PANCREAS							
Adenoma: islet cell	1	1	0	1	0	0.9123	1.0000
Carcinoma: islet cell	1	0	0	1	1	0.4093	0.7682
PARATHYROID GLAND							
Adenocarcinoma	0	0	0	0	1	0.3704	0.5769
Adenoma	0	1	0	0	0	1.0000	1.0000
PITUITARY							
Carcinoma: pars distalis	1	4	3	3	0	0.9829	1.0000
SKIN MISCELLANEOUS							
Sarcoma: squamous cell	1	0	0	0	0	1.0000	1.0000
SPINAL CORD CERVIC							
Malignant astrocytoma	0	0	0	0	1	0.2893	0.5147
SPLEEN							
Hemangioma	0	1	0	0	0	1.0000	1.0000
STOMACH							
ECL cell tumour	1	0	0	0	0	1.0000	1.0000
Systemic							
Hemangioma	0	1	0	1	0	0.8227	1.0000
Hemangioma/-sarcoma	0	1	1	1	0	0.7938	1.0000
Hemangiosarcoma	0	0	1	0	0	0.5593	
THYROID							
Adenoma: C-cell	2	4	3	7	2	0.8997	0.9673
Adenoma: follicular cell	0	2	0	1	0	0.9389	1.0000
Carcinoma: follicular cel	0	0	0	1	0	0.3390	
TONGUE							
Papilloma: squamous cell	0	0	0	0	1	0.2893	0.5147
UTERUS							
Adenocarcinoma: endometri	0	1	1	0	0	0.9273	1.0000
Adenoma: endometrial	0	0	0	2	0	0.5565	
Benign granular cell tumo	0	0	0	1	0	0.5372	
Fibroma	0	0	1	0	0	0.5854	
Fibrosarcoma	0	0	0	0	1	0.2253	0.3851
Hemangioma	0	0	0	1	0	0.5372	
Leiomyoma	2	0	0	0	0	1.0000	1.0000
Polyp: endometrial stroma	6	3	14	4	7	0.9010	0.7103
Sarcoma: endometrial stro	1	0	1	1	0	0.7278	1.0000

Appendix 4. References

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 Biopharmaceutical Statistics*. 8(1), 1-15.

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.

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.

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

/s/

Steven Thomson
6/19/2007 01:13:38 PM
BIOMETRICS

Karl Lin
6/19/2007 02:55:35 PM
BIOMETRICS
Concur with review