CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

22-015

STATISTICAL REVIEW(S)

STATISTICAL REVIEW AND EVALUATION
CLINICAL STUDIES

NDA/Serial Number:	22015
Drug Name:	MiraLax (Polyethylene Glycol 3350, NF Powder for Solution)
Indication(s):	Treatment of occasional constipation in patients with history of constipation.
Applicant:	Braintree Laboratories, Inc.
Date(s):	Received on December 8, 2005
Review Priority:	Standard
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Wen-Jen Chen, Ph.D. (HFD-725)
Concurring Reviewers:	Stella Grosser, Ph.D. (HFD-725)
Medical Division:	Gastroenterology Products
Medical Reviewer:	Kristin Buck, MD.
Project Manager:	Mr. Keith Olin
Keywords:	NDA review, clinical studies.

TABLE OF CONTENTS

1.0 EXECUTIVE SUMMARY OF STATISTICAL FINDINGS

1.1 Conclusions and Recommendations

From the statistical perspective, based upon the primary and secondary endpoint analyses, the two studies (851-CR1 and 851-ZCC) reviewed for this application provide substantial evidence to demonstrate that MiraLax is superior to placebo in treatment of occasional constipation in patients with history of constipation.

1.2 Brief Overview of Clinical Studies

The applicant submitted three studies (851-CR1, 851-ZCC, and 851-CR3) to support the use of MiraLax in the treatment of occasional constipation in patients with history of constipation. However, Study 851-CR3 was a single arm study to evaluate the safety of extended (1 year) use of MiraLax laxative in constipation patients; no placebo arm was included for the efficacy comparison. Accordingly, Study 851-CR3 is not further reviewed. In this review, the other two studies (851-CR1 and 851-ZCC) are the focus. The three clinical studies (851-CR1, 851-CR3, and 851-ZCC) are summarized in Table 1.2.1.

Table 1.2.1 Summary of Clinical Studies

STUDY	STUDY DESIGN	STUDY DRUG	OBJECTIVE	TREATMENT DURATION
851-CR1	Randomized Parallel; Double-Blinded.	MiraLax 17g/day vs. Placebo	Efficacy and Safety	6 months
851-CR3	Single arm; open-label	MiraLax 17g/day	Safety Chronic use	12 months
851-ZCC	Randomized Parallel; open-label.	MiraLax 17g/day vs. Zelnorm 6 mg BID	Safety and Efficacy	1 month

Study 851-CR1 was a six month, phase III, randomized, double blind, parallel, placebo-controlled, multi-center study to evaluate extended (6 month) use of MiraLax as compared to placebo in the treatment of patients with occasional constipation. Three hundred and six (306) healthy-constipated outpatients were enrolled according to objective constipation criteria (ROME) and were randomized to receive either MiraLax treatment or placebo in a parallel study design. The two treatment groups were compared for bowel movement (BM) frequency, ease of passage and straining, etc. However, the primary efficacy endpoint was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder).

Study 851-ZCC was a phase III, randomized, open label, parallel, active-controlled, multi-center study to evaluate 28-day use of MiraLax as compared to Zelnorm in the treatment of patients with occasional constipation. Patients were randomly assigned in a 1:1 ratio within each participating site. Two hundred thirty-nine (239) male and female patients that met a definition of constipation (ROME) and all other entry criteria were enrolled and randomized to receive either MiraLax or Zelnorm. Similar to Study 851-CR1, the primary efficacy endpoint was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder).

1.3 Statistical Issues and Findings

1.3.1 Study 851-CR1

- For the primary endpoint, the applicant's analysis indicated that at the end of six month study-period, a highly statistically significant 41% difference in treatment response between MiraLax and placebo was observed. In addition, this reviewer's analysis of the efficacy treatment comparison assessed by responder rate within each site does not find any single site dominates the superiority result of MiraLax to placebo.
- For the efficacy analysis assessed by the monthly-responder (defined as monthly treatment success), the applicant indicated that MiraLax treatment resulted in a much more rapid increase in the percent of patients successfully treated within the first month of therapy when compared with that of placebo (47% versus 9%). In addition, this reviewer's analysis on the treatment efficacy comparison assessed by the proportion of patients successfully treated (primary efficacy assessment) on each week shows that at week 2, a statistically significant difference between MiraLax and placebo in percent of patients successfully treated was observed.
- Finally, the secondary endpoint analyses performed by the applicant on number of successful weeks, number of successful weeks assessed by individual ROME symptoms, number of bowel movements, and global assessment all showed that MiraLax was superior to placebo.

Accordingly, based upon the efficacy analyses performed by the applicant and this reviewer on the primary and secondary endpoints, data provided by the applicant for this study demonstrated that the efficacy of MiraLax used in the treatment of occasional constipation in patients with history of constipation is superior to that of placebo.

1.3.2 Study 851-ZCC

- For the primary endpoint, the applicant's analysis indicated that the percentage of responder for MiraLax is significantly higher than that of Zelnorm using ITT population. In addition, this reviewer's analysis on the efficacy treatment comparison assessed by responder within each site suggests that no one site dominates the superiority result of MiraLax to Zelnorm.
- For the primary efficacy analysis assessed by the difference in proportions of successfully treated patients on each week (proportion of week-success) between the MiraLax and Zelnorm groups, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients over the four weeks of therapy. In addition, at week 2, the proportion of week-success for MiraLax (46%) is significantly higher than that of Zelnorm (27%).
- For the secondary endpoints, the applicant's analysis also demonstrated that at significance level of 0.05, the number of successful weeks for MiraLax was significantly higher than that of Zelnorm assessed by primary definition, ROME constipation

definition, and super-week definition (super-efficacy is defined as not satisfying any ROME criteria with no rescue medication).

As a result, based upon the analyses for the primary efficacy assessments (responder and proportion of week-success) and the secondary endpoints, Study 851-ZCC supports the claim that the efficacy of MiraLax is superior to placebo in treatment of occasional constipation in patients with history of constipation.

2.0 INTRODUCTION

2.1 Overview

In the section of "Clinical Overview" of the clinical study report, the applicant made the following observations with regard to MiraLax:

> MiraLax is a laxative composed of polyethylene glycol 3350 (PEG 3350). Because PEG is not metabolized or significantly absorbed, it remains in the lumen of the gastrointestinal tract where it exerts an osmotic effect. The osmotic activity of PEG thus increases the water content of stool with a resulting increase in stool volume.
>
> MiraLaX laxative was derived from Braintree Laboratories' PEG-electrolyte lavage solutions which were designed to cleanse the gut prior to diagnostic examination by rapidly inducing a voluminous liquid stool. The PEG lavages were formulated with electrolytes to prevent net gain or loss of electrolytes from the resulting diarrhea.

The applicant submitted three studies (851-CR1, 851-ZCC, and 851-CR3) to support the use of MiraLax in the treatment of occasional constipation in patients with history of constipation. However, Study 851-CR3 was a single arm study to evaluate the safety of extended (1 year) use of MiraLax laxative in constipation patients; no placebo arm was included for the efficacy comparison. Accordingly, Study 851-CR3 is not further reviewed. In this review, the other two studies (851-CR1 and 851-ZCC) are the focus. The three clinical studies (851-CR1, 851-CR3, and 851-ZCC) are summarized in Table 2.1.1.

Table 2.1.1 Summary of Clinical Studies

STUDY	STUDY DESIGN	STUDY DRUG	OBJECTIVE	TREATMENT DURATION
851-CR1	Randomized Parallel; Double-Blinded.	MiraLax 17g/day vs. Placebo	Efficacy and Safety	6 months
851-CR3	Single arm; open-label	MiraLax 17g/day	Safety Chronic use	12 months
851-ZCC	Randomized Parallel; open-label.	MiraLax 17g/day vs. Zelnorm 6 mg BID	Safety and Efficacy	1 month

Study 851-CR1 was a six month, phase III, randomized, double blind, parallel, placebo-controlled, multi-center study to evaluate extended (6 month) use of MiraLax as compared to placebo in the treatment of patients with occasional constipation. Three hundred and six (306) healthy-constipated outpatients were enrolled according to objective constipation criteria (ROME) and were randomized to receive either MiraLax treatment or placebo in a parallel study

design. The two treatment groups were compared for bowel movement (BM) frequency, ease of passage and straining, etc. However, the primary efficacy endpoint was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder).

Study 851-ZCC was a phase III, randomized, open label, parallel, active-controlled, multi-center study to evaluate 28-day use of MiraLax as compared to Zelnorm in the treatment of patients with occasional constipation. Patients were randomly assigned in a 1:1 ratio within each participating site. Two hundred thirty-nine (239) male and female patients that met a definition of constipation (ROME) and all other entry criteria were enrolled and randomized to receive either MiraLax or Zelnorm. Similar to Study 851-CR1, the primary efficacy endpoint was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder).

2.2 Data Sources

To assess the clinical efficacy of MiraLax in the treatment of occasional constipation in patients with history of constipation, this reviewer reviewed the NDA submission, dated December 8, 2006. In addition, data used by this reviewer's statistical analysis was submitted by the applicant on May 10, 2006 and located at "\\CDSESUB1\N22015\N_000\2006-05-10".

3.0 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study 851-CR1

Study Design and Endpoints

The objective of this study was to evaluate the safety and efficacy of extended (6 month) use of MiraLax laxative as compared to placebo in constipated patients, including a subgroup of elderly patients.

This was a six month, phase III, randomized, double blind, parallel, placebo-controlled, multi-center study to evaluate extended (6 month) use of MiraLax as compared to placebo in the treatment of patients with occasional constipation. Three hundred and six (306) healthy-constipated outpatients were enrolled according to objective constipation criteria (ROME) and randomized to receive either MiraLax treatment or placebo in a parallel study design. The two treatment groups were compared for bowel movement (BM) frequency, ease of passage and straining, etc.

Patients were allowed the use of bisacodyl 5mg tablets as rescue medication and were instructed to take 10mg of bisacodyl if they experienced severe discomfort due to their constipation, or if they had not had a BM in 4 days.

Male and female patients, who met the protocol definition of constipation but generally in good health, were enrolled. Of these patients, about 100 were expected to be 65 years of age or older. Enrolled study patients were instructed to stop all laxative treatments for a 14 day observation period. 306 eligible patients who met the study definition of constipation were randomized in a 2:1 ratio (Miralax to placebo) to a treatment group by a computer generated randomization scheme. The randomization schedule at each site was constructed using random sized blocks of 3 balanced treatment assignments in order to insure the specified 2:1 treatment ratio. Patients that met eligibility criteria at each site were sequentially assigned a kit number from the randomization schedule provided by Braintree.

Patients called into an IVRS (Interactive Voice Response System) each day to report their BM experiences for that day and answer questions related to the study efficacy and safety criteria. No safety, data monitoring or special steering or evaluation committees were formed or met during the study period. No interim analysis was performed.

The primary efficacy endpoint was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder). First, a treatment success-week (primary efficacy assessment) was defined as:

i) Three or more satisfactory stools per week, and
ii) one or fewer of the following ROME based criteria
    a. Straining in more than 25% of defecations
    b. Lumpy or hard stools in more than 25% of defecations
    c. Sensation of incomplete evacuation in more than 25% of defecations.

Then, the overall treatment success was further defined as a 0.50 or greater rate of successful treatment weeks versus total treatment weeks. In other words, a successfully treated patient had to have at least 50% of their treatment weeks scored as "successful".

The secondary efficacy endpoints included the following:

1) ROME Definition: A successful week was defined as not satisfying any 3 of 4 ROME constipation symptom criteria without the aid of rescue medication or prohibited laxative;
2) Super Efficacy: A successful week was defined as not satisfying any of the four ROME constipation symptom criteria without the aid of rescue medication or prohibited laxative;
3) A successful treatment week rate was also defined in terms of each individual ROME constipation symptom. A successful week was defined as not satisfying that constipation criterion without the aid of rescue medication or prohibited laxative.
4) A successful treatment week rate was also defined in terms of no use of rescue medication or prohibited laxative. A non-successful treatment week was any week for which either rescue medication or prohibited laxative was used.

Statistical Methodologies

The primary analysis was based upon an intent-to-treat (ITT) population which included all patients randomized and receiving any treatment. The primary efficacy analysis was on the

primary efficacy endpoint of overall treatment success or failure rate determined for each patient.

The analysis for the primary efficacy endpoint used the Cochran-Mantel-Haenszel (CMH) statistic stratified by site with no covariate adjustment to compare the treatment difference. Exact p-value was used for this comparison. In addition, a 95% confidence interval for the difference in proportions was also obtained for the non-stratified population.

Sites that recruited fewer than 24 intent-to-treat (ITT) patients were pooled to form larger pseudo sites in order to maintain at least 24 ITT patients for each site in the CMH stratified analysis. To meet this requirement, pseudo sites were created by pooling individual sites within a pre-determined geographic region. The specifics of this pooling algorithm were defined prior to un-blinding the study data and included in a detailed statistical analysis plan.

Secondary efficacy endpoints defined in terms of successful treatment rates were analyzed using analysis of variance (ANOVA) with factors for treatment group, pooled site, and interaction between treatment group and pooled site. Selected secondary endpoints were also analyzed using survival analysis to evaluate time to event. The time to treatment response was defined as time since first dose until obtaining response criteria. The duration of response was defined as the time of first obtaining the response criteria until the first time of failure to obtain the response criteria. The differences in response curves for the two treatment groups were compared using a log rank test. The estimated time to event and the proportion of patients obtaining the event at 4, 8, 12, 16, and 24 weeks were based on the Kaplan-Meier product limit method.

The sample size calculation was based upon the normal approximation to the binomial distribution. Using the results from a previous study and taking into account potential laxative use, the overall treatment success for the placebo group was expected to be approximately 40%. An absolute increase of 20 percentage points in overall treatment success with MiraLax over placebo (40% to 60%) was considered a clinically important improvement. Assuming a 40% placebo response rate for overall treatment success, based on a two-sided chi-squared test, a study size of 300 patients (200 on MiraLax and 100 on placebo) was expected to have 90% power to detect a treatment difference of 20% at the two-sided significance level of 0.05.

Patient Disposition

This study was conducted at 50 centers. Six hundred and nine (609) patients were screened and 306 patients were enrolled. Of the 306 patients, one patient was randomized in error by study personnel and did not receive study medication. Another patient was dropped from the study immediately following randomization due to complete non-compliance with study requirements. These two patients were removed from the Intent-to-Treat (ITT) analysis. Accordingly, 304 patients (including 75 with ages of 65 years or older) received study medication and were included in the ITT analysis.

Of the 304 ITT patients, 170 patients completed all 6 months of study. The reasons for discontinuation are given below in Table 3.1.1.1.

Table 3.1.1.1 (Applicant's) Reasons for Patient Discontinuation

	MiraLAX (n)	Placebo (n)
Completing Patients	62.3% (127)	43.0% (43)
Patients Discontinued	37.7% (77)	57.0% (57)
Reasons:		
Patient withdrew consent	18% (14)	25% (14)
Lack of efficacy	31% (24)	46% (26)
Non-compliance	12% (9)	12% (7)
Lost to follow-up	14% (11)	5% (3)
Adverse event	25% (19)	12% (7)

The applicant further indicated that patient withdrawals associated with Miralax or placebo treatment were proportionally equivalent for each reason category with the exception of withdrawals attributed to lack of efficacy. In this category, more than twice as many patients withdrew due to lack of efficacy associated with placebo treatment as did patients from MiraLax treatment.

Three hundred and three (303) patients did not meet study inclusion/exclusion criteria or otherwise failed screening during the 14 day washout period. The reasons for screen failures are given below in Table 3.1.1.2.

Table 3.1.1.2 (Applicant's) Reasons for Screen Failure

Patients Failed Screen	303
Reasons:	
Failed BM criteria	38% (115)
Failed inclusion criteria	24% (73)
Withdrew consent	13% (40)
Adverse event	1% (1)
Non-compliance	24% (74)

Demographics and Baseline Characteristics

In the study population, the majority of enrollees were female (258 or 85%). Forty-six males were enrolled. The applicant indicated that this gender disparity is consistent with previous constipation studies and with the overall demographics of constipation. The study population demographics are summarized in Table 3.1.1.3, below.

Table 3.1.1.3 (Applicant's) Study Demographics

	MiraLAX			Placebo			
	All	**Younger (<65 y)**	**Elderly (≥65 y)**	**All**	**Younger (<65 y)**	**Elderly (≥65 y)**	**p[1]**
Age (years)[2]							
n	204	153	51	100	76	24	0.46
Mean (SD)	53.1 (14.8)	46.6 (10.5)	72.7 (6.5)	54.4 (15.0)	48.4 (11.5)	73.5 (6.4)	
Gender							
Female	175 (86%)	144 (94%)	31 (61%)	83 (83%)	70 (92%)	13 (54%)	0.56
Male	29 (14%)	9 (6%)	20 (39%)	17 (17%)	6 (8%)	11 (46%)	
Race							
Caucasian	168 (82%)	122 (80%)	46 (90%)	87 (87%)	63 (83%)	24 (100%)	0.81
A. Am.	28 (14%)	25 (16%)	3 (6%)	11 (11%)	11 (14%)	0	
Other	4 (2%)	2 (1%)	2 (4%)	1 (1%)	1 (1%)	0	
Missing	4 (2%)	4 (3%)	0	1 (1%)	1 (1%)	0	
Ethnicity							
Hispanic	12 (5.9)	12 (8%)	0	7 (7%)	6 (8%)	1 (4%)	0.75
Non Hispanic	192 (94.1)	141 (92%)	51 (100%)	93 (93%)	70 (92%)	23 (96%)	
Weight (kg)							
Mean (SD)	74.7(16.3)	74.5(17.5)	75.1 (12.2)	75.1 (15.6)	73.4 (14.7)	80.3 (17.2)	0.65
Constipation Hx (yrs)	23.4 (18.7)	21.1 (15.8)	30.2 (24.4)	22.6 (19.2)	20.4 (16.2)	29.5 (25.8)	0.66
Mean (SD)							

(1) P-Value from CMH test controlling with pooled site for the categorical variables and from an ANOVA with terms for pooled site and treatment for the continuous variables.

(2) Age is calculated using of date of birth and screening visit (Visit 1) date.

SD = standard deviation; kg = kilograms; A. Am. = African American

Based upon Table 3.1.1.3, the applicant indicated that the treatment groups were similar with respect to age, racial distribution, weight, and constipation history. The average age of study participants was about 53 years, ranging in age from 20 to 92 years of age. About 84% of study enrollees were Caucasian and 13% were African American, reflecting national racial population distribution. Study patients weighed an average of about 75 kg. There were no demographic related statistically significant differences between the treatment groups.

Applicant's Efficacy Analysis Results and Conclusions

Primary endpoint analysis

The primary efficacy endpoint for treatment response was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder). Table 3.1.1.4 presented the primary responder analysis using ITT patient population.

Table 3.1.1.4 (Applicant's) Primary efficacy responder analysis using ITT population

Responder[1]	MiraLAX n (%)	Placebo n (%)	All	95% CI[2]	p[3]
All Patients (n)	204	100	304		
Yes	106 (52%)	11 (11%)	117 (39%)	31.8, 50.2	<0.001
No	98 (48%)	89 (89%)	187 (61%)		
Elderly (≥65 y)	51	24	75		
Yes	30 (59%)	3 (13%)	33 (44%)	27.4, 65.2	<0.001
No	21 (41%)	21 (87%)	42 (56%)		
Non-Elderly (<65y)	153	76	229		
Yes	76 (50%)	8 (11%)	84 (37%)	28.6, 49.7	<0.001
No	77 (50%)	68 (89%)	145 (63%)		
Males	29	17	46		
Yes	13 (45%)	1 (6%)	14 (30%)	17.7, 60.2	0.007
No	16 (55%)	16 (94%)	32 (70%)		
Females	175	83	258		
Yes	93 (53%)	10 (12%)	103 (40%)	30.9, 51.3	<0.001
No	82 (47%)	73 (88%)	155 (60%)		
Caucasian	172	88	260		
Yes	89 (52%)	10 (11%)	99 (38%)	30.4, 50.4	<0.001
No	83 (48%)	78 (89%)	161 (62%)		
Non-Caucasian	32	12	44		
Yes	17 (53%)	1 (8%)	18 (41%)	21.5, 68.1	0.014
No	15 (47%)	11 (92%)	26 (59%)		

(1) A successful treatment week is defined as >=3 satisfactory bowel movements, with 1 or no additional ROME symptom criteria, and without the aid of rescue medication or prohibited laxative during the week; a responder must have at least a 0.50 rate of successful treatment weeks (based on number of actual treatment weeks). Days with missing data are not included in computing success. A patient with fewer than 8 weeks of data will be counted as a failure.

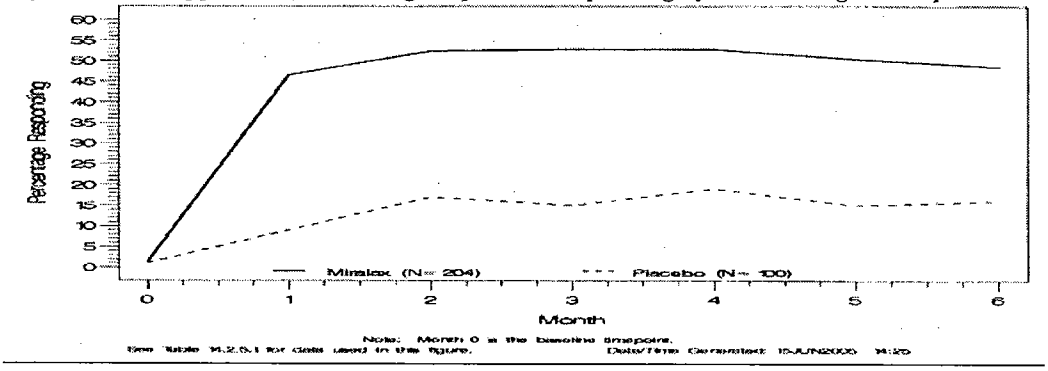
(2) Confidence interval (CI) for the difference between MiraLax and Placebo is from a Cochran-Mantel-Haenzsel test or Fisher's Exact Test (for race).

(3) P-value for the difference between MiraLax and Placebo is from a pooled site stratified Cochran-Mantel-Haenzsel test or Fisher's Exact Test (for race).

Based upon results from Table 3.1.1.4, the applicant indicated that the primary responder analysis using ITT population showed a highly statistically significant 41% difference in treatment response between MiraLax and placebo ($p<0.001$). The elderly subpopulation achieved similar efficacy (46%). More over, the percentage of patients who responded successfully was more than 4 times higher with MiraLax than with placebo, regardless of age, gender, or race.

In addition, the proportions of successfully treated patients (monthly-responder), according to the primary efficacy definition, for each month of the study for both treatments were displayed by Figure 3.1.1.1.

Figure 3.1.1.1 (Applicant's) Percentage of patients responding by month using ITT Population



Based upon Figure 3.1.1.1, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients within the first month of therapy. The maximum response occurred by the second month and the response then remained fairly level thereafter. However, the response to placebo was much less dramatic (about 9% in the first month) and remained at a low level over the course of the study.

Secondary endpoint analysis

For the secondary efficacy endpoints assessed by ROME definition and "super efficacy" for each treatment week, Table 3.1.1.5 presented the analysis results for the number of successful treatment weeks according to each definition.

Table 3.1.1.5 (Applicant's) Secondary endpoint analysis for number of successful weeks using ITT population

Responder Definition	MiraLAX	Placebo	All	p[1]
Mean Treatment Weeks	19.5	15.4	-	-
Primary Definition[2]	196	95	291	
Mean (SD)	12.0 (9.8)	3.4 (5.8)	9.2 (9.6)	<0.001
% of weeks	61.4%	21.8%		
ROME[3]	202	100	302	
Mean (SD)	12.9 (10.0)	3.8 (6.2)	9.9 (9.9)	<0.001
% of weeks	66.2%	24.4%		
Super Efficacy[4] (n)	196	95	291	
Mean (SD)	9.2 (9.0)	2.2 (4.7)	6.9 (8.5)	<0.001
% of weeks	47.3%	14.4%		

(1) P-Value from an ANOVA with terms for treatment, pooled-site, and treatment by pooled-site interaction.
(2) >3 satisfactory bowel movements, with 1 or no additional ROME symptom criteria, and without the aid of rescue medication or prohibited laxative during the week.
(3) ROME definition not met without aid of rescue medication.
(4) No ROME symptom criteria met, without aid of rescue medication. SD = Standard Deviation.

Based upon Table 3.1.1.5, the Applicant indicated that MiraLax treated patients had about 4 times as many successful treatment weeks as placebo patients by any definition. As might be expected, there were fewer successful "super efficacy" treatment weeks for both therapies due to the more strict definition that a successful treatment week could have none of the four individual ROME symptom criteria or use of rescue laxative. However, even by this rigorous definition, the applicant indicated that nearly 50% of treatment weeks were successful for MiraLax patiants versus 14% for placebo.

In addition, Table 3.1.1.6 demonstrated the analysis results for the number of successful weeks assessed by the four individual ROME components.

Table 3.1.1.6 (Applicant's) Number of successful weeks assessed by individual ROME symptom using ITT population

Responder Definition[2]	MiraLAX	Placebo	All	p[1]
Treatment Weeks	19.5	15.4	-	-
ROME #1				
< 3 Satis. BM (n)	202	100	302	<0.001
Mean (SD)	13.5 (9.8)	5.6 (7.4)	10.9 (9.8)	
% of weeks	68.9%	36.4%		
ROME #2				
Strain >25% (n)	202	100	302	<0.001
Mean (SD)	12.4 (9.9)	3.1 (5.6)	9.3 (9.8)	
% of weeks	63.6%	20.1%		
ROME # 3				
Hard Stool >25%	202	100	302	<0.001
Mean (SD)	14.3 (10.1)	4.5 (6.8)	11.1 (10.2)	
% of weeks	73.3%	29.2%		
ROME #4				
Incomplete >25%	202	100	302	<0.001
Mean (SD)	10.6 (9.2)	4.3 (6.7)	8.5 (8.9)	
% of weeks	54.4%	27.9%		

(1) P-Value from an ANOVA with terms for treatment, pooled-site, and treatment by pooled-site interaction.
(2) Specific ROME symptom not met, without aid of rescue medication.

Table 3.1.1.6 showed the number of successful treatment weeks for each of the four ROME symptom criteria. The table entries indicated weeks where the ROME constipation symptom was not met (i.e. a successful treatment week). The differences between MiraLax and placebo in individual ROME symptoms were all statistically significant with the most dramatic differences occurring in straining (symptom 2) and hard stool (symptom 3).

Table 3.1.1.7 displayed other secondary endpoint analyses assessed by the number of bowel movements, global assessment, and rescue medication use.

Table 3.1.1.7 (Applicant's) Other secondary endpoint analyses using ITT population

Responder Definition	MiraLAX	Placebo	All	p[1]
Mean BM/wk (n)	202	100	302	
Mean (SD)	7.9 (4.5)	5.6 (5.5)	7.1 (5.0)	<0.001
Mean Satisfactory BM/wk (n)	202	100	302	<0.001
Mean (SD)	5.4 (3.6)	2.7 (2.1)	4.5 (3.4)	
Mean CSBM/wk[2]	202	100	302	
Mean (SD)	5.0 (4.2)	2.1 (2.7)	4.0 (4.0)	<0.001
Global Assess.[3]	202	100	302	
Mean weeks (SD)	12.5 (8.9)	5.2 (7.1)	10.1 (9.0)	<0.001
% weeks[4]	64.1%	33.8%		
Rescue Med Use	198	97	295	
Mean tabs/wk (SD)	2.8 (6.0)	3.9 (7.1)	3.2 (6.4)	0.138

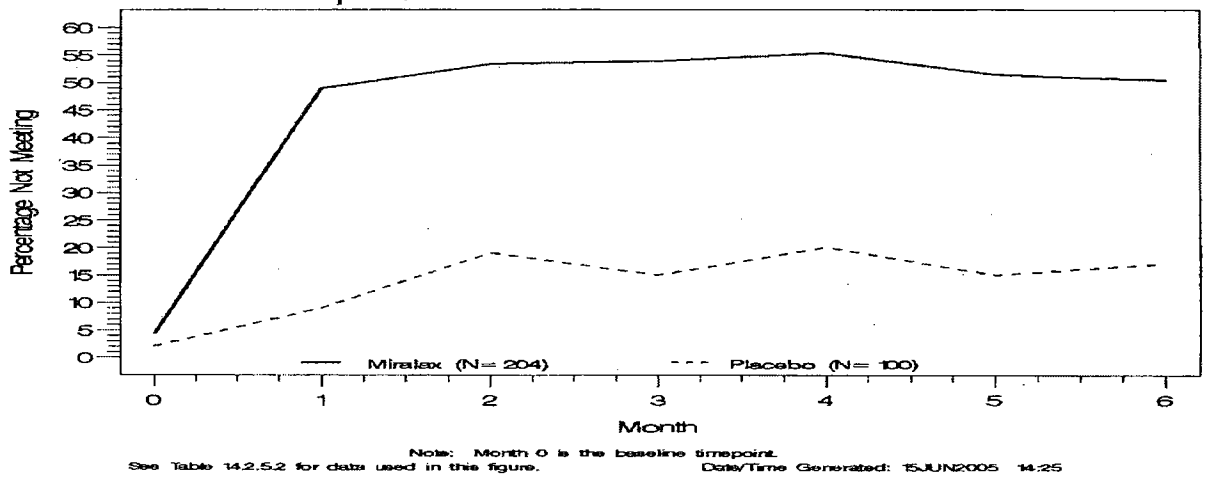
(1) P -Value from an ANOVA with terms for treatment, pooled-site, and treatment by pooled-site interaction.
(2) Complete, Spontaneous BM, without aid of rescue medication.
(3) Number of weeks that patients indicated that they had adequate relief.
(4) Mean number of MiraLax treatment weeks = 19.5; mean placebo weeks = 15.4. SD = Standard Deviation

The applicant indicated that Table 3.1.1.7 showed statistically significant differences between treatments in the total number of bowel movements (BM) per week as well as the number of "Satisfactory BM" per week. MiraLax patients achieved one bowel movement per day. In fact, MiraLax treatment resulted in nearly double the number of "Satisfactory BM" per week (about 5.4) as compared to placebo (about 2.7). This level of weekly "Satisfactory BM" output for placebo met the study definition of constipation (fewer than 3 satisfactory BM's per week). MiraLax also performed much better than placebo in an analysis for "Complete, Spontaneous BM" (CSBM). In this analysis, a successful CSBM was defined as a patient score for a BM as "complete" and occurring on a day in which no stimulant rescue laxative was taken.

For the "Global Assessment of Efficacy" (GEA), patients taking MiraLax noted that 64% of their treatment weeks were satisfactory as compared to 34% of placebo treatment weeks. The 30% difference was statistically significant. However, for the number of tablets used per week, although MiraLax-treated patients on average used fewer tablets of the rescue medication, this difference did not reach statistical significance.

Finally, Figure 3.1.1.2 displayed the proportion (as percent) of successfully treated patients assessed by the ROME definition for each month of the study for both treatments. The treatment success was defined as a 0.50 or greater rate of successful treatment weeks versus total treatment weeks.

Figure 3.1.1.2 (Applicant's) Percentage of patients not meeting Rome definition of constipation using ITT Population



Based upon Figure 3.1.1.2, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients assessed by ROME criteria within the first month of therapy. The maximum response occurred by the second month and the response then remained fairly level thereafter. The response to placebo was much less dramatic in the first month and remained at low levels over the course of the study.

Statistical Reviewer's Comments and Analysis

In order to validate the applicant's efficacy claim, this reviewer first performs the following two analyses: 1) efficacy comparison by site based upon the primary endpoint (responder) and 2) percent of patients successfully treated by treatment group and week. Then, this reviewer comments on the efficacy of MiraLax demonstrated by the study.

Reviewer's Analysis

1) Efficacy comparison assessed by responder

In order to explore whether the efficacy of MiraLax to placebo assessed by the primary endpoint responder was dominated by any sites, this reviewer analyzes the differences in proportions with regard to the primary endpoint by site to compare the efficacy between two treatments using MITT population. The sites used in this analysis are the sites provided by the data set submitted by the applicant. Table 3.1.1.8 presents the result.

Table 3.1.1.8 (Reviewer's) Responder rate by treatment group and site using MITT population

SITE NUMBER	MIRALAX % (n/N)	PLACEBO % (n/N	SITE NUMBER	MIRALAX % (n/N	PLACEBO % (n/N)	SITE NUMBER	MIRALAX % (n/N)	PLACEBO % (n/N)
Site 101	60.0 (3/5)	0.0 (0/2)	Site 118	40.0 (2/5)	0.0 (0/2)	Site 134	25.0 (1/4)	0.0 (0/3)
Site 102	30.0 (3/10)	0.0 (0/5)	Site 119	60.0 (6/10)	25.0 (1/4)	Site 135	75.0 (6/8)	0.0 (0/4)
Site 103	0.0 (0/2)	0.0 (0/1)	Site 120	40.0 (2/5)	0.0 (0/2)	Site 136	33.3(2/6)	0.0 (0/3)
Site 104	33.3(1/3)	0.0 (0/1)	Site 121	50.0 (5/10)	25.0 (1/4)	Site 137	40.0 (2/5)	0.0 (0/2)
Site 105	50.0 (1/2)	50.0 (1/2)	Site 122	100.0 (3/3)	0.0 (0/1)	Site 139	50.0 (1/2)	0.0 (0/1)
Site 107	0.0 (0/2)	0.0 (0/2)	Site 123	100.0 (2/2)	0.0 (0/1)	Site 141	50.0 (3/6)	0.0 (0/3)
Site 108	0.0 (0/1)	0.0 (0/1)	Site 124	20.0 (1/5)	0.0 (0/2)	Site 142	33.3(1/3)	50.0 (1/2)
Site 109	No data	0.0 (0/1)	Site 125	0.0 (0/4)	50.0 (1/2)	Site 143	80.0 (4/5)	0.0 (0/3)
Site 110	50.0 (3/6)	0.0 (0/3)	Site 126	66.7 (2/3)	0.0 (0/1)	Site 144	80.0 (4/5)	0.0 (0/2)
Site 112	57.0 (4/7)	33.3(1/3)	Site 127	50.0 (1/2)	No data	Site 145	100.0 (5/5)	50.0 (1/2)
Site 113	No data	0.0 (0/1)	Site 128	0.0 (0/2)	0.0 (0/1)	Site 146	85.7 (6/7)	50.0 (2/4)
Site 114	40.0 (4/10)	0.0 (0/4)	Site 129	83.3 (5/6)	50.0 (1/2)	Site 147	100.0 (1/1)	0.0 (0/1)
Site 115	60.0 (3/5)	0.0 (0/3)	Site 130	75.0 (3/4)	0.0 (0/3)	Site 148	50.0 (1/2)	0.0 (0/2)
Site 116	75.0 (6/8)	0.0 (0/4)	Site 131	25.0 (1/4)	0.0 (0/1)	Site 149	60.0 (6/10)	0.0 (0/4)
Site 117	20.0 (1/5)	33.3(1/3)	Site 132	0.0 (0/2)	0.0 (0/1)	Overall	52.0 (106/204)	11.0 (11/100)

Based upon the results from Table 3.1.1.8, since the largest site (site 102) only had 15 patients included in the MITT population and of the 44 sites, only eight sites with patients greater than 10 (18%), basically, this was a small site study. In addition, for the most of sites, the percents of responders in the MiraLax group are much greater than that in the Placebo group. Accordingly, one may conclude that no particular large site was found to dominate the superiority result of MiraLax to Placebo assessed by the primary endpoint (responder - overall treatment success at end of study duration).

2) Efficacy comparison assessed by treatment success week

In order to explore the efficacy of MiraLax at week 2 after drug administration, this reviewer compares the proportions of patients successfully treated (primary efficacy assessment) between MiraLax and placebo by week. Figure 3.1.1.1 displays the results.

Figure 3.1.1.1 (Reviewer's) Percentage of patients successfully treated compared between treatment groups at each study week using MITT population

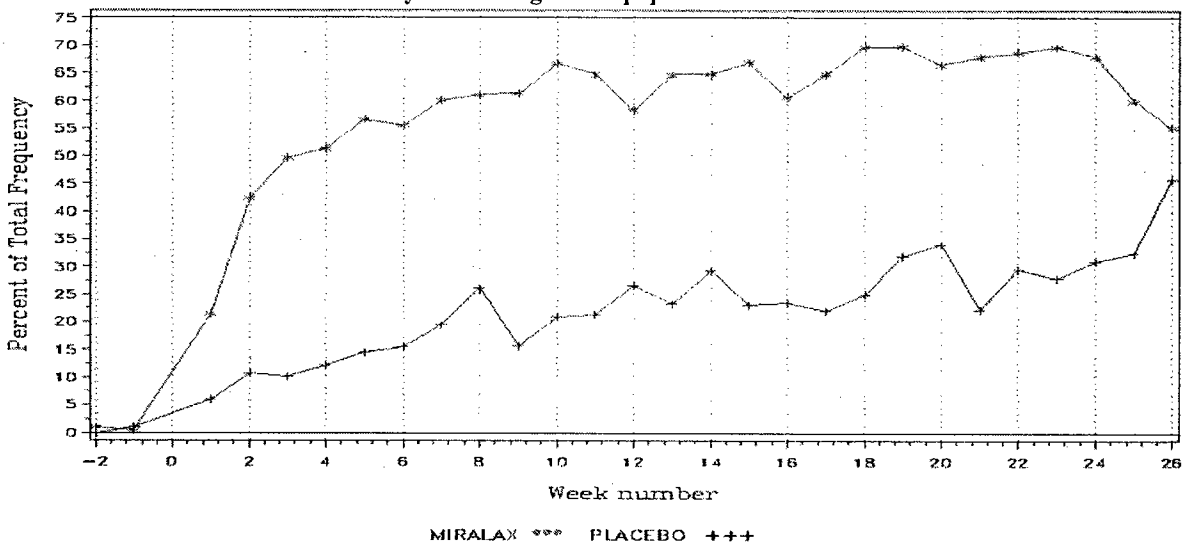


Figure 3.1.1.1 indicates that the percent of patients successfully treated in the MiraLax group rapidly increases from baseline week (0.5%) to week 2 (42%) while that in the placebo group increases from 1% to 11%. The 31% difference on the percent of patients successfully treated between MiraLax and Placebo at week 2 is significant ($p<0.001$) and supports the efficacy of MiraLax for the use of two weeks.

Reviewer's Comments on the efficacy of MiraLax

For the primary endpoint of responder (overall treatment success), the applicant's analysis indicated that at end of six month study-period, a highly statistically significant 41% difference in treatment response between MiraLax and placebo was observed. In addition, this reviewer's analysis on the efficacy treatment comparison assessed by responder within each site does not find a particular large site to dominate the superiority result of MiraLax to Placebo assessed by the primary endpoint (responder).

For the applicant's primary efficacy analysis assessed by the monthly-responder, MiraLax treatment resulted in a much more rapid increase in the percent of patients successfully treated within the first month of therapy (monthly-responder) when compared with that of placebo (47% versus 9%). In addition, this reviewer's analysis on the treatment efficacy comparison assessed by the proportion of patients successfully treated (primary efficacy assessment) on each week

shows that at week 2, a highly statistically significant 31% difference between MiraLax and placebo in percent of patients successfully treated at week 2 was observed.

Finally, the secondary endpoint analyses performed by the applicant on number of successful weeks, number of successful weeks assessed by individual ROME symptom, number of bowel movements, and global assessment all showed that MiraLax was superior to placebo.
Accordingly, based upon the efficacy analyses performed by the applicant and this reviewer on the primary and secondary endpoints, data provided by the applicant for this study demonstrated that the efficacy of MiraLax used in the treatment of occasional constipation in patients with history of constipation is superior to that of placebo.

3.1.2 Study 851-ZCC

Study Design and Endpoints

The objective of this study was to evaluate the safety and efficacy of use of MiraLax laxative as compared to Zelnorm in patients with constipation.

This was a phase III, randomized, open label, parallel, active-controlled, multi-center study to evaluate 28-day use of MiraLax as compared to Zelnorm in the treatment of patients with occasional constipation. Patients were randomly assigned in a 1:1 ratio to receive MiraLax or Zelnorm within each participating site. The randomization schedule was generated by [redacted] using SAS version 8.2. The randomization schedule at each site was constructed using random blocks of 2 patients, which provided balanced treatment assignments in order to insure the specified 1:1 treatment ratio. Patients that met study eligibility criteria (including the constipation definition) at each site were sequentially assigned a kit number from the randomization schedule provided by Braintree.

Two hundred thirty-nine (239) male and female patients that met a definition of constipation (ROME) and all other entry criteria were enrolled and randomized to either MiraLax or Zelnorm by a computer generated randomization scheme. Two patients were randomized in error by study personnel and did not receive study medication. These two patients were not included in the Intent-to-Treat (ITT) population. Of the 237 ITT patients, 31 were 65 years of age or older. Randomized patients were treated with study medication each day for up to 28 days.

Patients were also allowed the use of bisacodyl 5mg tablets as rescue medication and were instructed to take 10mg of bisacodyl if they experienced severe discomfort due to their constipation, or if they had not had a BM in 4 days.

Patients called into an IVRS (Interactive Voice Response System) each day to report their BM experiences for that day. Following input of the patient identifiers and security code, the IVRS prompted the patients to answer questions related to the study efficacy and safety criteria. No interim analysis was performed.

The primary efficacy endpoint was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder). First, a treatment success week was defined as no use of rescue laxative and met the following criteria:

i.) Satisfactory stool greater or equal to 3 per week;

ii.) One or fewer of the following additional ROME based criteria
   a) straining in more than 25% of defecations;
   b) lumpy or hard stools in more than 25% of defecations; and
   c) sensation of incomplete evacuation in more than 25% of defecations.

Then, similar to Study 851-CR1, the overall treatment success (responder) was further defined as a 0.50 or greater rate of successful treatment weeks versus total treatment weeks. In other words, a successfully treated patient had to have at least 50% of their treatment weeks scored as "successful".

Secondary efficacy endpoints included the following:

1) ROME Definition: a successful week was defined as not satisfying any 3 of 4 ROME criteria without the aid of rescue medication or prohibitive laxative. Only days in which data have been reported counted toward the endpoint calculation. The rate of successful treatment weeks was defined in the same manner as for the primary endpoint.
2) Super efficacy: A successful week was defined as not satisfying any of the 4 ROME criteria without the aid of rescue medication or prohibitive laxative. The rate of successful treatment weeks is defined the same as for the primary endpoint.
3) For each ROME criterion, a successful week was defined as satisfying that criterion without the aid of rescue medication or prohibited laxative.
4) A successful treatment week was also defined in terms of no use of rescue medication or prohibitive laxative. A non-successful treatment week was any week for which either rescue medication or prohibitive laxative was used, etc.

Statistical Methodologies

The primary analysis group was based upon an intent-to-treat (ITT) analysis and included all patients randomized and receiving any treatment. The primary efficacy analysis was based on the primary efficacy endpoint of overall treatment success or failure determined for each patient. The null hypothesis H_0 is: "There is no difference in the proportion of responders between MiraLax and Zelnorm" versus the alternative hypothesis H_a "There is a difference in the proportion of responders between MiraLax and Zelnorm."

The primary analysis for the between treatment comparison used the Cochran-Mantel- Haenszel statistic stratified by site with no covariate adjustment. The difference was the weighted difference of responder rates between the MiraLax group and the Zelnorm group. The weight for each site was proportional to the number of patients in each treatment group. Sites that recruited less than 20 ITT patients were pooled to form larger pseudo sites in order to maintain at least 20 ITT patients for each site in the Cochran- Mantel-Haenszel (CMH) stratified analysis. To meet

this requirement, pseudo sites were created by pooling individual sites within a pre-determined geographic region. The specifics of this pooling algorithm were defined prior to un-blinding the study data and included in a detailed statistical analysis plan.

Secondary efficacy endpoints defined in terms of successful treatment rates were analyzed using analysis of variance with factors for treatment group, pooled-site, and interaction between treatment group and pooled-site. Treatment emergent adverse event rates were descriptively presented by body system, preferred term, severity, and relationship to treatment for each treatment group. Differences in adverse event rates between treatment groups were assessed using Fishers Exact Test.

The sample size calculation was based upon the normal approximation to the binomial distribution. Using the results from a previous study in which Zelnorm was compared to placebo in female patients with constipation predominant IBS, and taking into account potential laxative use, the overall treatment success rate for the Zelnorm group was expected to be approximately 40%. An absolute increase of 20 percentage points in overall treatment success with MiraLax over Zelnorm (60% to 40%) was considered a clinically important improvement. Assuming a 40% Zelnorm response rate for overall treatment success, based on a two-sided chi-squared test, a study size of 240 patients (120 MiraLax and 120 Zelnorm) will have 80% power to detect a treatment difference of 20% at the two-sided significance level of 0.05.

Patient Disposition

This study was conducted at 25 centers. Two hundred and thirty seven (237) patients (including 31 elderly) were enrolled and received treatment. Two hundred three (203) patients completed all 4 weeks of study. The reasons for discontinuation are given below in Table 3.1.2.1.

Table 3.1.2.1 (Applicant's) Reasons for Patient Discontinuation

	MiraLAX n (%)	Zelnorm n (%)
Patients Discontinued	14 (11.7%)	20 (17.1%)
Reasons:		
Patient withdrew consent	6 (5.0%)	7 (6.0%)
Lack of efficacy	1 (0.8%)	1 (0.9%)
Non-compliance	4 (3.3%)	4 (3.4%)
Lost to follow-up	3 (2.5%)	3 (2.6%)
Adverse event	0	5 (4.3%)
Completing Patients	106 (88.3%)	97 (82.9%)

Demographics and Baseline Characteristics

In the study population, the applicant indicated that the majority of enrollees were female (213). Twenty-four males were enrolled. This gender disparity is consistent with previous constipation studies and with the overall demographics of constipation. In addition, male patients were specifically excluded by protocol amendment 3. The average age of study participants was 46

years, ranging in age from 19 to 81 years of age. Elderly patients were specifically excluded following the approval of protocol amendment 3.

The average duration of constipation reported by all patients was 17.5 years. About 64% of study enrollees were Caucasian, 24% were African American, and 13% were of Hispanic or Latino ethnicity. The percentage of African American patients is higher than the national average, which can be attributed to the geographic location of study centers. Study patients weighed an average of about 76 kg. There were no demographic related, statistically significant differences between the treatment groups. The study population demographics were summarized in Table 3.1.2.2.

Table 3.1.2.2 (Applicant's) Study Demographics

	MiraLAX		Zelnorm		
	All	**Younger (<65 y)**	**All**	**Younger (<65 y)**	**P[1]**
Age (years)[2]					
n	120	103	117	103	0.75
Mean(SD)	46.1 (14.4)	42.2 (11.5)	46.9 (14.5)	43.5 (11.7)	
Gender					
Female	109 (91%)	96 (93%)	104 (89%)	93 (90%)	0.59
Male	11 (9%)	7 (7%)	13 (11%)	10 (10%)	
Race					
Caucasian	72 (60%)	57 (55%)	79 (68%)	66 (64%)	0.766
A. Am.	31 (26%)	29 (28%)	26 (22%)	25 (24%)	
Other	5 (4%)	5 (5%)	5 (4%)	5 (5%)	
Missing	12 (10%)	12 (12%)	7 (6%)	7 (7%)	
Ethnicity					
Hispanic	18 (15%)	18 (17%)	13 (11%)	13 (13%)	0.328
Non-Hispanic	102 (85%)	85 (83%)	103(88%)	89 (86%)	
Missing	0	0	1 (1%)	1 (1%)	
Weight (kg)					
Mean (SD)	77.0 (22.4)	77.8 (23.6)	75.8 (18.3)	75.4 (18.4)	0.68
Constipation Hx (years)					
	16.2 (14.2)	15.4 (13.6)	18.9 (18.2)	16.7 (15.6)	0.27

(1) P-Value from CMH test controlling for pooled-site for the categorical variables, and from an ANOVA with terms for pooled-site and treatment for the continuous variables.
(2) Age is calculated using the date of birth and screening visit date.
Note: SD = standard deviation; kg = kilograms; A. Am. = African American.

Applicant's Efficacy Analysis Results and Conclusions

Primary endpoint analysis

The primary efficacy endpoint for treatment response was assessed on the basis of a binary outcome of overall treatment success (responder) or failure (non-responder). Table 3.1.2.3 presented the primary responder analysis using ITT patient population.

Table 3.1.2.3 (Applicant's) Primary efficacy responder analysis using ITT population

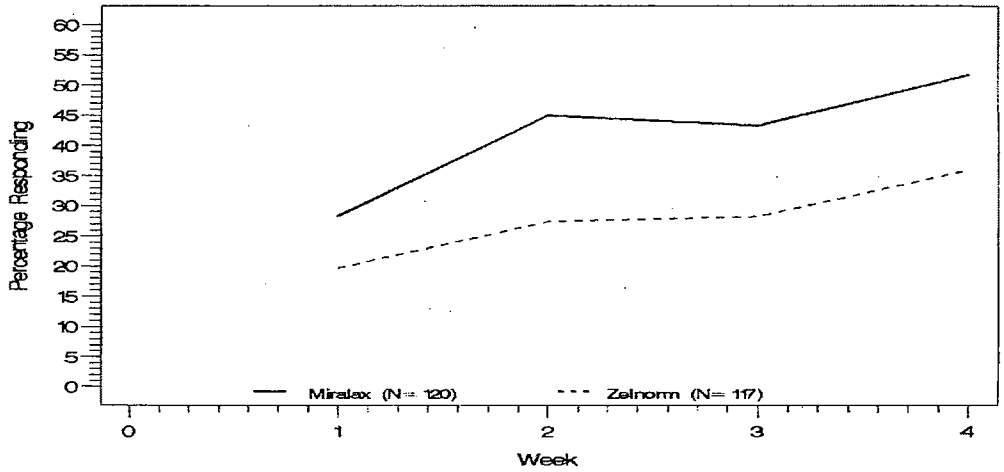
Responder	**MiraLAX n (%)**	**Zelnorm n (%)**	**95% CI[1]**	**P[2]**
All Patients (n)	120	117		
Yes	60 (50.0%)	36 (30.8%)	7.0, 31.5	**0.003**
No	60 (50.0%)	81 (69.2%)		
Non Elderly (<65 y)	103	103		
Yes	49 (47.6%)	33 (32.0%)	2.3, 28.7	**0.032**
No	54 (52.4%)	70 (68.0%)		

(1) Confidence interval (CI) for the difference between MiraLax and Zelnorm is from a Cochran-Mantel-Haenszel test.
(2) P-value for the difference between MiraLax and Zelnorm is from a pooled-site stratified Cochran-Mantel-Haenszel test.

As shown in Table 3.1.2.4, the applicant indicated that the primary responder analysis using ITT population showed a highly statistically significant difference (19.2 % with p = 0.003) in treatment response between MiraLax and Zelnorm. When analyzing only the non-elderly patients for which Zelnorm now has FDA approved labeling, the statistically significant difference in response favoring MiraLax remains (p = 0.032).

In addition, for each week of the study, the proportions of successfully treated patients (according to the primary efficacy definition) were displayed by Figure 3.1.2.1.

Figure 3.1.2.1 (Applicant's) Percentage of patients successfully treated by week using ITT Population



Note: Percent of patients responding to therapy by week for the primary efficacy measure. A successful treatment week was defined as > 3 satisfactory bowel movements with no more than 1 additional ROME symptom criteria without the aid of rescue medication or prohibited laxative. At Weeks 2 through 4, the difference was statistically significant (p=0.005, 0.015, 0.015, respectively).

Based upon Figure 3.1.2.1, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients over the four weeks of therapy (to 50%). The difference in proportions of successfully treated patients on each week between the MiraLax and Zelnorm groups reached statistical significance by Week 2 and remained significant through Week 4.

Secondary endpoint analysis

For the secondary efficacy endpoints assessed by ROME definition and "super efficacy" (defined as not satisfying any of the 4 ROME criteria without the aid of rescue medication or prohibitive laxative) for each treatment week, Table 3.1.2.4 presented the analysis results for the number of successful treatment weeks according to each definition.

Table 3.1.2.4 (Applicant's) Secondary endpoint analysis for number of successful weeks using ITT population

Responder Definition	MiraLAX n=120	Zelnorm n=117	p[1]
Primary Definition[2] Mean (SD)	1.79 (1.51)	1.19 (1.36)	**0.003**
ROME[3] Mean (SD)	1.84 (1.53)	1.28 (1.35)	**0.006**
Super Efficacy[4] Mean (SD)	1.09 (1.35)	0.71 (1.12)	**0.028**

(1) P-Value from an ANOVA with terms for treatment, pooled-site, and treatment by pooled-site interaction.
(2) >3 satisfactory bowel movements, with 1 or no additional ROME symptom criteria, and without the aid of rescue medication or prohibited laxative during the week.
(3) ROME constipation definition not met without aid of rescue medication.
(4) No ROME constipation symptom criteria met, without aid of rescue medication. SD = Standard Deviation.

Based upon Table 3.1.2.4, the applicant indicated that the number of successful weeks when applying the primary responder definition was highly statistically significant in favor of MiraLax. This response persisted when both groups were analyzed using the clinically accepted ROME Definition. As might be expected, there were fewer successful Super Efficacy treatment weeks for both therapies due to the more strict definition which required that a successful treatment week could have none of the four individual ROME constipation symptom criteria. However, even by this rigorous definition, the number of successful treatment weeks was found to be statistically significant in favor of MiraLax.

In addition, Table 3.1.2.6 demonstrated the analysis for the number of successful weeks assessed by each of the four individual ROME symptom criteria.

Table 3.1.2.6 (Applicant's) Number of successful weeks assessed by individual ROME symptom using ITT population

Responder Definition[2]	MiraLAX N=120	Zelnorm N=117	p[1]
ROME #1 ≥ 3 Satis. BM Mean (SD)	2.43 (1.6)	2.39 (1.5)	0.703
ROME #2 Strain <25% Mean (SD)	1.78 (1.6)	1.37 (1.4)	0.065
ROME # 3 Hard Stool <25% Mean (SD)	2.13 (1.5)	1.47 (1.4)	**0.001**
ROME #4 Incomplete <25% Mean (SD)	1.37 (1.5)	1.23 (1.4)	0.448

(1) P-Value from an ANOVA with terms for treatment, pooled-site, and treatment by pooled-site interaction.
(2) Specific ROME symptom not met, without aid of rescue medication.

The applicant indicated that as shown in Table 3.1.2.6, the difference in stool consistency (lumpy/hard stools) between MiraLax and Zelnorm was highly statistically significant (P= 0.001). The differences in BM frequency, incomplete evacuation, and straining all favored MiraLax, however none was statistically significant.

Table 3.1.2.7 displayed other secondary endpoint analyses assessed by the number of bowel movements, global assessment, and rescue medication use.

Table 3.1.2.7 (Applicant's) Other secondary endpoint analyses using ITT population

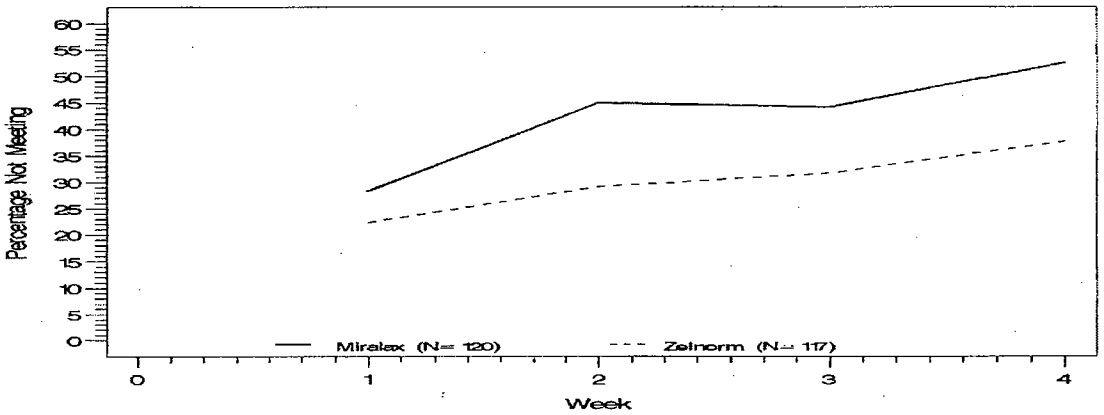
Responder Definition	MiraLAX	Zelnorm	p[1]
Mean BM/wk (n)	118	116	
Mean (SD)	10.42 (7.7)	8.48 (4.9)	**0.019**
Mean Satisfactory BM/wk (n)	118	116	0.072
Mean (SD)	7.09 (5.7)	5.84 (4.3)	
Mean CSBM/wk[2]	118	116	
Mean (SD)	5.56 (5.2)	4.80 (4.2)	0.162
Global Assess.[3]	118	116	
Mean weeks (SD)	1.95 (1.4)	1.63 (1.3)	0.081
Rescue Med Use	102	93	
Mean tabs/wk (SD)	1.40 (3.4)	1.00 (2.3)	0.268

(1) P-Value &om an ANOVA with terms for treatment, pooled-site, and treatment by pooled-site interaction.
(2) Complete, Spontaneous BM, without aid of rescue medication.
(3) Number of weeks that patients indicated that they had adequate relief.

The applicant indicated that Table 3.1.2.7 showed a statistically significant difference between treatments in the total number of bowel movements (BM) per week. All other BM related measures favored MiraLax, however these were not statistically significant. In addition, the number of rescue medication tablets used per week was not significantly different between treatment groups.

Finally, Figure 3.1.2.2 displayed the proportion of successfully treated patients assessed by the ROME Definition for each week of the study for both treatments.

Figure 3.1.2.2 (Applicant's) Percentage of patients not meeting Rome definition of constipation using ITT Population



Note: Percent of patients not meeting ROME definition by week. According to this definition, a successful treatment week was defined as a patient reporting no more than 1 ROME symptom criterion, without the aid of rescue medication or prohibited laxative. The ROME symptoms for constipation are:
[a] < 3 satisfactory bowel movements per week ; [b] Lumpy or hard stools in more than 25% of defecations;
[c] Straining in more than 25% of defecations; [d] Sensation of incomplete evacuation in more than 25% of defecations.
At Weeks 2 through 4, the difference was statistically signi6cant (p=0.011, 0.047, C.022, respectively)

Based upon Figure 3.1.2.2, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients during the four week treatment period when assessed by not meeting Rome definition of constipation using ITT population. A statistically significant difference favoring MiraLax was achieved at Week 2 and persisted through Week 4.

Statistical Reviewer's Comments and Analysis

In order to validate the applicant's efficacy claim, this reviewer first performs the efficacy comparison for MiraLax versus Zelnorm by site based upon the primary endpoint (responder). Then, this reviewer comments on the efficacy of MiraLax demonstrated by the study.

Reviewer's Analysis

In order to explore whether the efficacy of MiraLax to Zelnorm assessed by the primary endpoint (responder) was dominated by any sites, this reviewer analyzes the differences in proportions with regard to the primary endpoint by site to compare the efficacy between two treatments (MiraLax and Zelnorm) using MITT population. The sites used in this analysis are the original sites recorded in the data set submitted by the applicant. Table 3.1.2.8 presents the result.

Table 3.1.2.8 (Reviewer's) Responder rate by treatment group and site using MITT population

SITE NUMBER	MIRALAX % (n/N)	ZELNORM % (n/N	SITE NUMBER	MIRALAX % (n/N	ZELNORM % (n/N	SITE NUMBER	MIRALAX % (n/N)	ZELNORM % (n/N
Site 102	50.0 (4/8)	11.0 (1/9)	Site 121	0.0 (0/4)	25.0(1/4)	Site 148	0.0 (0/2)	0.0 (0/2)
Site 107	80.0 (4/5)	33.3 (2/6)	Site 124	66.70 (2/3)	33.3 (1/3)	Site 149	0.0 (0/3)	33.30 (1/3)
Site 112	0.0 (0/2)	0.0 (0/3)	Site 129	50.0 (2/4)	0.0 (0/4)	Site 151	62.50 (10/16)	33.30 (5/15)
Site 114	75.0 (3/4)	25.0 (1/4)	Site 135	66.70 (4/6)	33.3 (2/6)	Site 152	0.0 (0/1)	50.0 (1/2)
Site 115	0.0 (0/4)	100.0 (3/3)	Site 136	50.0 (3/6)	16.70 (1/6)	Site 153	50.0 (4/8)	12.50 (1/8)
Site 116	33.3 (2/6)	16.70 (1/6)	Site 141	66.70 (8/12)	41.70 (5/12)	Site 155	42.90 (3/7)	83.33 (5/6)
Site 117	60.0 (3/5)	25.0 (1/4)	Site 142	0.0 (0/2)	No Data			
Site 119	100.0 (2/2)	50.0 (1/2)	Site 144	100.0 (4/4)	25.0 (1/4)			
Site 120	25.0 (1/4)	50.0 (2/4)	Site 146	50.0 (1/2)	0.0 (0/1)	Overall	50.0 (60/120)	31.0 (36/117)

Based upon the results from Table 3.1.2.8, it is noted that the responder rate of MiraLax for the largest site (site 151) is 29.2% higher than that of Zelnorm. However, of the 9 sites with number of patients greater 10, six sites show that the responder rates of MiraLax are higher than that of Zelnorm more than 29%. In addition, since of the total 24 sites, only 9 sites have number of patients greater than 10, basically, like Study851-CR1, this was also a small site study. Finally, for most small sites, the responder rate of MiraLax is found to be much higher than that of Zelnorm. As a result, one may conclude that no particular large site is found to dominate the superiority result of MiraLax to Zelnorm assessed by the primary endpoint (responder - overall treatment success at end of study duration).

Reviewer's Comments on the efficacy of MiraLax

For the primary endpoint (responder), the applicant's analysis indicated that the percentage of responders for MiraLax is significantly higher (p = 0.003) than that of Zelnorm using ITT population. In addition, this reviewer's analysis on the efficacy treatment comparison assessed by responders within each site suggests that no single site dominates the superiority result of MiraLax to Zelnorm assessed by the primary endpoint responder

For the primary efficacy analysis assessed by the difference in proportions of successfully treated patients in each week (proportion of week-success) between the MiraLax and Zelnorm groups, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients over the four weeks of therapy (approximate 50%). In addition, at week 2, the proportion of patients with week-success treated by MiraLax, based upon Figure 3.1.2.1, is significantly higher than that of patients treated by Zelnorm (46% vs. 27%, P=0.005).

For the secondary endpoints, the applicant's analysis also demonstrated that at significance level of 0.05, the number of successful weeks for MiraLax was significantly higher than that of Zelnorm assessed by primary definition, ROME constipation definition, and super-week definition.

As a result, based upon the analyses for the primary efficacy assessments (responder and proportion of week-success) and the secondary endpoints, Study 851-ZCC supports that the efficacy of MiraLax is superior to placebo in treatment of occasional constipation in patients with history of constipation.

3.2 Evaluation of Safety

3.2.1 Study 851-CR1

The applicant indicated that except for gastrointestinal adverse effects, no treatment emergent adverse effect differences were detected when long term MiraLax therapy was compared to placebo, even in the elderly. No significant gender or race related effects on adverse events were observed and no substantive differences were observed for patients taking narrow therapeutic index medicines or for high risk patients.

The only statistically significant difference detected (versus placebo) was gastrointestinal disorder in the general population for MiraLax. This difference appeared to be associated with slightly more abdominal symptoms, diarrhea, loose stools, flatulence and nausea in association with MiraLax treatment, although individually these differences were not statistically significant. Most of these reports were mild or moderate in severity and did not result in electrolyte abnormalities. These effects are considered to be consistent with the mode of action of a laxative.

3.2.2 Study 851-ZCC

The applicant indicated MiraLax and Zelnorm presented a similar adverse event profile, with the only significant difference being an increase in Nervous System Disorders (headache, dizziness) in Zelnorm patients. No age or gender related effects were observed. Gastrointestinal events (diarrhea and nausea) occurred with the greatest frequency. Most of these reports were mild or moderate in severity and these effects are consistent with the mode of action of a laxative. No Deaths or other Serious Adverse Events occurred during this trial or during 1 month following discontinuation of treatment.

Finally, the applicant concluded that overall, adverse events experienced by both treatment groups proved to be consistent with laxative use, and their approved labeling. In particular, the expected gastrointestinal events of diarrhea, nausea, and flatulence were not significantly different between groups.

4.0 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 GENDER, RACE, AND AGE

In order to assess the consistency of the treatment effect of MiraLax versus placebo (Study 851-CR1) or MiraLax versus Zelnorm (Study 851-ZCC) across subgroups, this reviewer performed the subgroup analysis by Mantel-Haenszel test for the primary endpoint responder (overall treatment success) using MITT patient population. The subgroups analyzed for Studies 851-CR1 and 851-ZCC are Gender (Male and Female), Race (White versus Non-White), and Age group (age $\leq$ 65 and age > 65).

4.1.1 Study 851-CR1

Gender group (Female versus Male)

Table 4.1.1.1 presents the results of treatment efficacy comparisons for the MiraLax to placebo by Gender group (Female versus Male).

Table 4.1.1.1 (Reviewer's) Proportion difference analysis on overall success using MITTpopulation

	MIRALAX m/n (%)	PLACEBO m/n (%)	P-VALUE†
Male	13/29 (45.0)	1/17 (6.0)	0.006
Female	93/175 (53.0)	10/83 (12.0)	< 0.0001

†: **P-value for Mantel-Haenszel test;**

Table 4.1.1.1 shows that for both Male and Female sub-groups, the percentages of overall success for MiraLax are significantly higher than that of placebo (p=0.006 for Males p < 0.0001 for females).

Race group (White versus Non-White)

Table 4.1.1.2 presents the results of treatment efficacy comparisons for the MiraLax to placebo by Race group (White versus Non-White).

Table 4.1.1.2 (Reviewer's) Proportion difference analysis on overall success using MITTpopulation

	MIRALAX m/n (%)	PLACEBO m/n (%)	P-VALUE†
White	86/168 (52.0)	10/87 (12.0)	< 0.0001
Non-White	20/36 (56.0)	1/13 (8.0)	0.0031

†: P-value for Mantel-Haenszel test;

Table 4.1.1.2 shows that for both White and Non-White sub-groups, the percentages of overall success for MiraLax are significantly higher than that of placebo ($p < 0.0001$ for White and $p=0.0031$ for Non-White).

Age group (age ≤ 65 and age > 65)

Table 4.1.1.3 presents the results of treatment efficacy comparisons for the MiraLax to placebo by Age group (age ≤ 65 and age > 65).

Table 4.1.1.3 (Reviewer's) Proportion difference analysis on overall success using MITTpopulation

	MIRALAX m/n (%)	PLACEBO m/n (%)	P-VALUE†
Age > 65	29/50 (58.0)	3/23 (13.0)	0.0004
Age ≤ 65	77/154 (50.0)	8/77 (10.0)	< 0.0001

†: P-value for Mantel-Haenszel test;

Table 4.1.1.3 shows that for both age sub-groups (age ≤ 65 and age > 65), the percentages of overall success for MiraLax are significantly higher than that of placebo ($p=0.0004$ for age > 65 and $p < 0.0001$ for age ≤ 65).

4.1.2 Study 851-ZCC

Gender group (Female versus Male)

Table 4.1.2.1 presents the results of treatment efficacy comparisons for the MiraLax to Zelnorm by Gender group (Female versus Male).

Table 4.1.2.1 (Reviewer's) Proportion difference analysis on overall success using MITTpopulation

	MIRALAX m/n (%)	ZELNORM m/n (%)	P-VALUE†
Male	8/11 (73.0)	4/13 (31.0)	0.045
Female	52/109 (48.0)	32/104 (31.0)	0.012

†: P-value for Mantel-Haenszel test;

Table 4.1.2.1 shows that for both Male and Female sub-groups, the percentages of overall success for MiraLax are significantly higher than that of Zelnorm (p=0.045 for Males p = 0.012 for females).

Race group (White versus Non-White)

Table 4.1.2.2 presents the results of treatment efficacy comparisons for the MiraLax to Zelnorm by Race group (White versus Non-White).

Table 4.1.2.2 (Reviewer's) Proportion difference analysis on overall success using MITTpopulation

	MIRALAX m/n (%)	ZELNORM m/n (%)	P-VALUE†
White	32/72 (44.0)	25/79 (32.0)	0.15
Non-White	28/48 (58.0)	11/38 (29.0)	0.007

†: P-value for Mantel-Haenszel test;

Table 4.1.2.2 shows that only for Non-White sub-group, the percentage of overall success for MiraLax is significantly higher than that of Zelnorm (p =0.007).

For White subgroup, the percentage of overall success for MiraLax is numerically higher than that of Zelnorm (44.0% for MiraLax versus 32.0% for Zelnorm).

Age group (age ≤ 65 and age > 65)

Table 4.1.2.3 presents the results of treatment efficacy comparisons for the MiraLax to Zelnorm by Age group (age ≤ 65 and age > 65).

Table 4.1.2.3 (Reviewer's) Proportion difference analysis on overall success using MITTpopulation

	MIRALAX m/n (%)	ZELNORM m/n (%)	P-VALUE†
Age > 65	11/17 (65.0)	2/13 (15.0)	0.008
Age ≤ 65	49/103 (48.0)	34/104 (33.0)	0.03

†: P-value for Mantel-Haenszel test;

Table 4.1.2.3 shows that for both age sub-groups (age ≤ 65 and age > 65), the percentages of overall success for MiraLax are significantly higher than that of Zelnorm (p=0.008 for age > 65 and p=0.03 for age ≤ 65).

4.2 Other Special/Subgroup Populations- Not applicable

5.0 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

5.1.1 Study 851-CR1

- For the primary endpoint, the applicant's analysis indicated that at the end of six month study-period, a highly statistically significant 41% difference in treatment response between MiraLax and placebo was observed. In addition, this reviewer's analysis of the efficacy treatment comparison assessed by responder rate within each site does not find any single site dominates the superiority result of MiraLax to placebo.
- For the efficacy analysis assessed by the monthly-responder (defined as monthly treatment success), the applicant indicated that MiraLax treatment resulted in a much more rapid increase in the percent of patients successfully treated within the first month of therapy when compared with that of placebo (47% versus 9%). In addition, this reviewer's analysis on the treatment efficacy comparison assessed by the proportion of patients successfully treated (primary efficacy assessment) on each week shows that at week 2, a statistically significant difference between MiraLax and placebo in percent of patients successfully treated was observed.
- Finally, the secondary endpoint analyses performed by the applicant on number of successful weeks, number of successful weeks assessed by individual ROME symptoms, number of bowel movements, and global assessment all showed that MiraLax was superior to placebo.

Accordingly, based upon the efficacy analyses performed by the applicant and this reviewer on the primary and secondary endpoints, data provided by the applicant for this study demonstrated that the efficacy of MiraLax used in the treatment of occasional constipation in patients with history of constipation is superior to that of placebo.

5.1.2 Study 851-ZCC

- For the primary endpoint, the applicant's analysis indicated that the percentage of responder for MiraLax is significantly higher than that of Zelnorm using ITT population. In addition, this reviewer's analysis on the efficacy treatment comparison assessed by responder within each site suggests that no one site dominates the superiority result of MiraLax to Zelnorm.
- For the primary efficacy analysis assessed by the difference in proportions of successfully treated patients on each week (proportion of week-success) between the MiraLax and Zelnorm groups, the applicant indicated that MiraLax treatment resulted in a rapid increase in the number of successfully treated patients over the four weeks of therapy. In addition, at week 2, the proportion of week-success for MiraLax (46%) is significantly higher than that of Zelnorm (27%).
- For the secondary endpoints, the applicant's analysis also demonstrated that at significance level of 0.05, the number of successful weeks for MiraLax was significantly higher than that of Zelnorm assessed by primary definition, ROME constipation

definition, and super-week definition (super-efficacy is defined as not satisfying any ROME criteria with no rescue medication).

As a result, based upon the analyses for the primary efficacy assessments (responder and proportion of week-success) and the secondary endpoints, Study 851-ZCC supports the claim that the efficacy of MiraLax is superior to placebo in treatment of occasional constipation in patients with history of constipation.

5.2 Conclusions and Recommendations

From the statistical perspective, based upon the primary and secondary endpoint analyses, the two studies (851-CR1 and 851-ZCC) reviewed for this application provide substantial evidence to demonstrate that MiraLax is superior to placebo in treatment of occasional constipation in patients with history of constipation.

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

/s/

Wen-Jen Chen
9/22/2006 05:25:25 PM
BIOMETRICS

Stella Grosser
9/25/2006 02:08:25 PM
BIOMETRICS

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:	22,015 / Serial 000
Drug Name:	MiraLax™ (Sitagliptin Phosphate)
Indication(s):	OTC Laxative for occasional constipation
Applicant:	Braintree Laboratories, Inc.
Date(s):	Submitted 12/16/05
Review Priority:	Standard
Biometrics Division:	Division 6, HFD-705
Statistical Reviewer:	Steve Thomson, HFD-705
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D., HFD-705
Medical Division:	Gastroenterology Products, HFD-180
Toxicologist:	Reviewer: Tamal Chakraborti, Ph.D., HFD-180 Team Leader: Jasti Choudary, Ph.D., HFD-180
Project Manager:	Keith Olin, HFD-180
Keywords:	Bayesian analysis, Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY

This submission was intended to assess the carcinogenic potential of daily administration of MiraLax™ when administered orally (by gavage) to mice and rats for a period of up to two years.

1.1. Conclusions and Recommendations

The submission reports on the results of two animal studies of carcinogenicity:

Study [redacted] 382018: A 24-Month Oral (Gavage) Carcinogenicity Study of MiraLax™ PEG-3350 in Mice,

and,

Study [redacted] 382009: A 24-Month Oral (Gavage) Carcinogenicity Study of MiraLax™ PEG-3350 in Rats.

In both studies there were four treatment groups (i.e., a control, and three nominal dosages of MiraLax: Control, 1.5, 3, and 6 g/kg/day), labeled as Control, Low, Medium, and High dose groups respectively. Vehicle was deoinized water. In both studies treatment was administered orally by gavage for up to 24 months. The Sponsor reports that in males of both species treatment was continued to 104 weeks. Due to mortality in the high dose treatment group in female mice, dosing was terminated at 94 weeks. Animals in the other female mice dosing groups were treated to the end of the study (104 weeks). Similarly, for female rats in the medium and high dose groups, dosing was stopped at 98 weeks. For controls and the low dose group, dosing in female rats was continued to 103 weeks.

In both genders in mice the high dose group generally had the highest mortality rate, while the low and medium dose groups were more or less similar in mortality, but with somewhat higher mortality than in the control group (please see Appendix 1 for details). However in mice the tests of homogeneity in survival were only clearly statistically significant in females (Males: Logrank p = 0.0062, Wilcoxon p = 0.0064, proportional hazards test of trend p = 0.0064). Differences were close to significance in male mice (Males: Logrank p = 0.0826, Wilcoxon p = 0.0506, trend p = 0.1745). Treatment group related differences in mortality were not apparent in rats (Male Rats: Logrank p = 0.5134 & Wilcoxon p = 0.5648, Female rats: Logrank p = 0.5495 & Wilcoxon p = 0.4559). Plots and some details are provided in Appendix 1. Results from a Bayesian analysis of survival were similar (please see Appendix 2).

For the tests for tumorigenicity, in both species, the only statistically significant results were in trend tests. That is, even without adjusting for multiplicity in testing, in both species and both genders there were no statistically significant pairwise differences between the control group and the high dose group. In mice, prior to adjusting for multiplicity, the tests of trend in

bronchio-alveolar adenoma in the lungs in males and benign skin pilomatricoma in females were statistically significant ($p \leq 0.0344$ and $p \leq 0.0345$, respectively). In rats, also prior to adjusting for multiplicity, the tests of trend in benign islet cell carcinoma in the pancreas of males and undifferentiated malignant carcinoma in the mammary gland of females were statistically significant or close to it ($p \leq 0.0432$ and $p \leq 0.0517$, respectively). However as discussed in Section 1.3.1, (1) the assumptions of the tests of trend are not clearly satisfied and thus the reported p-values may not be appropriate, and (2) even if appropriate, upon adjusting for multiplicity using the Haseman-Lin-Rahman rules, none would be considered statistically significant. So no tests of tumorigenicity would be considered to be statistically significant. It should be noted that absence of proof is not proof of absence. Nonetheless these results are consistent with the notion of no particular carcinogenic signal.

1.2. Brief Overview of the Studies

Two studies, both typical rodent studies, were submitted:

Study [redacted] -382018: A 24-Month Oral (Gavage) Carcinogenicity Study of MiraLax™ PEG-3350 in Mice,

and,

Study [redacted] -382009: A 24-Month Oral (Gavage) Carcinogenicity Study of MiraLax™ PEG-3350 in Rats.

In both studies there were four treatment groups,each starting with 65 animals (i.e., a control, and three nominal dosages of MiraLax: Control, 1.5, 3, and 6 g/kg/day), labeled as Control, Low, Medium, and High dose groups respectively. Vehicle was deoinized water. In both studies treatment was administered orally by gavage for up to 24 months. The Sponsor reports that in both studies, in males, treatment continued to 104 weeks. Due to mortality in the high dose treatment group in female mice, dosing was terminated at 94 weeks. Other female mice were treated to the end of the study (104 weeks). Similarly in the medium and high dose group in female rats dosing was stopped at 98 weeks. Otherwise, dosing in female rats was stopped at 103 weeks.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

Several issues, typical of such analyses, are considered in the following discussion. These include details of the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1. Survival Analysis:

Both logrank and Wilcoxon tests were used to test homogeneity of survival among the treatment groups, including the control group. Tests of dose related trend using a Cox proportional odds model were also performed. These involved testing multiple hypotheses, but from the point of view of finding differences among treatment groups (i.e., minimizing Type II error) would be conservative. Appendix 1 reviews the animal survival analyses in some detail. Appendix 2 provides an alternative Bayesian analysis of survival.

2. Tests in Neoplasms:

The Sponsor notes that in both studies for most organs animals at risk were only exhaustively microscopically examined in the control group and the the high dose group (6 g/kg/day). In the mouse study, in the low (1.5 g/kg/day) and medium (3 g/kg/day) dose groups generally only 40-50 selected animals were examined. In the rat study even fewer animals in these middle dose groups were microscopically examined. Unless either all animals at risk were examined or those animals to be examined were chosen at random independent of cues to tumor status, it is not clear how to adjust the analysis for the actual number of animals microscopically examined in the treatment group. Thus, except for a couple of organs, only the pairwise comparisons between the high dose group and the controls fully satisfy assumptions needed for the statistical tests. The Sponsor's approach is to only to report those pairwise tests, with the few trend tests where essentially all animals were analyzed. This is a reasonable approach.

However, since generally most animals were microscopically examined, the FDA analysis includes the test for trend as well as the pairwise comparisons between the high dose group and the control. Note this is equivalent to assuming that either no tumors would be found in those animals that were not examined by the pathologist or that the method of selecting animals is independent of their tumor status. Both assumptions are clearly false. Since this clearly causes a bias, in the tables displaying this analysis in Appendix 3, the p-values for thw tests of trend where not all anumals received a histopathological analysis are enclosed in parentheses, indicating they should only be used for rough guidance. The pairwise comparisons between the high dose group and control do not have this limitation and are not enclosed in parentheses. Finally, note that 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). The Peto tumorigenicity analyses were conducted using the FDA program supported by Dr. Ted Guo and others.

3. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involves a large number of statistical tests, which in turn necessitates 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 high dose group, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Based on simulations and their experience, Lin & Rahman (1998) proposed a 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 we will use the observed incidence in the pooled vehicle groups to decide if a tumor is rare or common. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

4. Validity of the Designs:

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. From the survival plots in the Appendix, it is evident that this value was clearly superceded in both studies, and in fact was exceeded by the number of animals that survived to the terminal sacrifice. This may indicate adequate exposure in both studies, but such a conclusion requires the expertise of the toxicologist.

Traditionally, in analyses performed in the United States, the highest dose should be close to the Maximum Tolerated Dose (MTD) to achieve the greatest likelihood of tumorigenicity. 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 Sponsor did not provide data sets for the animal weights. However, summary weight data was provided in the Sponsor's reports. The reported weight data for rats started at Week 26, and weights starting at this time are used instead of baseline for assessing the possible weight decrement in the high dose groups. Data for mice started at baseline. The table below indicates the mean weights at the indicated time points. In female mice dosing in the high dose group was terminated at 94 weeks. Similarly, in the medium and high dose group in female rats, dosing was stopped at 98 weeks. The entries for female mice and female rats include mean weights at

control groups. The mortality data for the rat study in Tables 7 and 8 indicate that there was no strong evidence of mortality differences among the treatment groups. Thus while the weight data may suggest that the MTD was exceeded at least for male rats, the survival data suggest the opposite problem (i.e., the high dose is under the MTD)..

The combination of the body weight gain data and the mortality information indicate that the high dose used in the mouse and the female rat studies are close to the MTD. The evaluation of the appropriateness of the high doae in the female rat study is inconclusive. However, the above evaluation of the appropriateness of the designs and whether or not the doses were sufficiently close to the MTD is based on some rules derived from data of 200 NCI carcinogen bioassays. Information regarding clinical signs and histopathological data, plus other possible considerations, are well beyond the expertise of this reviewer, but presumably would be used by the toxicologist in the final assessment of the adequacy of these experiments.

1.3.2. Statistical Findings

In both genders in mice the high dose group generally had the highest mortality rate, while the low and medium dose groups were more or less similar in mortality, but with somewhat higher mortality than in the control group (please see Appendix 1 for details). However in mice the tests of homogeneity in survival were only clearly statistically significant in females, consistent hypothesis of differences in survival. The results of the tests of homogeneity in survival in the different studies are displayed in Table 2. below:

Table 2. Tests of Homogeneity and Trend in Survival

	Mice			Rats		
Gender	Log Rank	Wilcoxon	Trend	Log Rank	Wilcoxon	Trend
Male	0.0826	0.0506	0.1745	0.5134	0.5648	0.1673
Female	0.0062	0.0064	0.0064	0.5495	0.4559	0.3047

For the tests for tumorigenicity, in both species, the only potentially statistically significant results were in trend tests. Even without adjusting for multiplicity, there were no statistically significant pairwise differences between the control group and the high dose group in either study. In mice, prior to adjusting for multiplicity the tests of trend in bronchio-alveola adenoma in the lungs males and benign skin pilomatricoma in females were statistically significant ($p \leq 0.0344$ and $p \leq 0.0345$, respectively). In rats, also prior to adjusting for multiplicity, the tests of trend in benign islet cell carcinoma in the pancreas of males and undifferentiates malignant carcinoma in the mammary gland of females were statistically significant or close to it ($p \leq 0.0432$ and $p \leq 0.0517$, respectively). The Haseman-Lin-Rahman rules indicate that to adjust for multiplicity, for an overall roughly 10% Type I error rate, common tumors should be tested at a 0.005 level and rare tumors at a 0.025 level. Using the incidence in the control group as a guide, only the first of these tumors would be classified as common. Then, as discussed in Section 1.3.1, (1) the assumptions of the tests of trend are not clearly satisfied, so the reported p-values may not be appropriate, and (2) even if appropriate,

upon adjusting for multiplicity using the Haseman-Lin-Rahman rules above, none would be considered statistically significant. So no tests of tumorigenicity would be considered to be statistically significant. Again, absence of proof is not proof of absence. Nonetheless these results are consistent with the notion of no particular carcinogenic signal

2. INTRODUCTION

2.1. Overview

Results from a study in [redacted]:CD-1® (ICR) BR mice and a study in [redacted]:CD®(SD)IGS BR rats were submitted to assess the carcinogenic potential of MiraLax.

2.2. Data Sources

For both studies, the following SAS transport data sets were included in a compact disk provided by the Sponsor:

382018FT, 38201MT, 38009FT, and 38009MT.

These data sets show the tumorigenicity results for the female mice, male mice, female rats, and male rats, respectively. No other data sets were provided.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

Results on both studies are presented below.

3.2.1. Study [redacted]-382018: A 24-Month Oral (Gavage) Carcinogenicity Study of MiraLax™ PEG-3350 in Mice,

There were four treatment groups (i.e., a vehicle control, and three nominal dosages of MiraLax: Control, 1.5, 3, and 6 g/kg/day), labeled as Control, Low, Medium, and High dose groups respectively. Vehicle was deionized water. Each treatment group initially had 65 [redacted]:CD-1®(ICR) BR mice, including 5 additional mice per group to accommodate accidental deaths in young mice. The Sponsor states that after four weeks of dosing, animals were selected at random to achieve a level of 60 animals per dose group. Treatment was administered orally by gavage daily for 104 weeks in all male dose groups and the the Control, Low, and Medium

dose groups in females. When survival reached the protocol specified level of 20 female mice at Week 94, dosing in the High dose group in females was stopped.

Males were 46 days old when received on 8 July and females were 44 days old when received on 6 May. Dosing for females was initiated on 20 May 2003, while dosing for males was initiated 23 July 2003. During the study animals were housed individually. Food and water were available ad libitum. The Sponsor states that detailed physical examinations were made on all animals each week. Body weights were recorded weekly, beginning approximately one week before initiation of dosing.

3.2.1.1 Sponsor's Results and Conclusions

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

Survival analysis:

The Sponsor provided the results of generalized Wilcoxon tests of homogeneity in survival across the four treatment groups (including the controls). When the overall test of homogeneity was rejected, the Sponsor provided the results of pairwise tests of each of the three MiraLax treatment groups with the control group. Note that differences among the four treatment groups were close to statistical significance in males, and quite statistically significant in female mice.

These mortality results are summarized in the following table, Table 3. For each treatment group, the number of animals who died of causes related to treatment, the number of animals at risk, and the Kaplan-Meier estimate of the percent who survived to that time point are presented. The p-value of the test labeled "Overall" is for the overall test of homogeneity over the four treatment groups. When this was statistically significant, the p-values of the pairwise tests of differences from control are presented.

Table 3. Summary of Mortality in Mice: Deaths/At Risk (KM)

Males Time Interval	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
0-52	2/60 (97%)	4/60 (93%)	5/60 (92%)	5/60 (92%)
53-78	5/57 (88%)	12/56 (73%)	12/53 (71%)	18/55 (62%)
79-92	15/52 (63%)	13/44 (52%)	10/41 (54%)	11/37 (43%)
93-EOS	9/37 (47%)	15/31 (26%)	12/31 (33%)	6/26 (33%)
Terminal Sacrifice	28	15	19	19
p-value	Overall: 0.0536	Not Tested	Not Tested	Not Tested

Overall test of Trend : p=0.0661

Table 3. (cont.) Summary of Mortality in Mice: Deaths/At Risk (KM)

Females Time Interval	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
0-52	2/60 (97%)	5/60 (92%)	3/60 (95%)	1/60 (98%)
53-78	10/57 (80%)	14/55 (68%)	11/56 (76%)	22/59 (62%)
79-92	9/47 (64%)	9/41 (53%)	13/44 (53%)	14/37 (38%)
93-EOS	9/38 (49%)	11/32 (35%)	8/30 (39%)	10/23 (22%)
Terminal Sacrifice	29	21	22	13
p-value	Overall: 0.0064*	vs Ctrl: 0. 0405*	vs Ctrl: 0.1996	Vs Ctrl: 0.0005*

Overall test of Trend : p=0.0014*

Tumorigenicity analysis:

The Sponsor conducted Peto type analyses to compare the incidence of various neoplasms (see Tables A.3.1 and A.3.2 in Appendix 3). Even without adjusting for the multiplicity of comparisons, neither any of the tests of pairwise differences in tumorgenicity, nor any of the few trend tests, were statistically significant.

3.2.1.2 FDA Reviewer's Results

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

Survival analysis:

As with the Sponsor's analysis, note that differences among the four treatment groups were close to statistical significance in males, and quite statistically significant in female mice. (Males: logrank p = 0.0826 and Wilcoxon p = 0.0506, Females: logrank p = 0.0062 and Wilcoxon p = 0.0064).

Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1, along with more details of the analysis. The following tables (Table 4 for male mice, Table 5 for female mice) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and give the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage is the percent survived at the end of the interval, as estimated using a Kaplan-Meier estimate on the ungrouped data. Note again the high dose group seems to have higher mortality.

Table 4. Summary of Male Mice Mortality: Number/At Risk (KM estimate)

Period (Weeks)	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
1-50	1/60 (98%)	4/60 (93%)	4/60 (93%)	5/60 (92%)
51-78	6/57 (88%)	12/56 (73%)	13/53 (71%)	18/55 (62%)
79-91	15/52 (63%)	12/44 (53%)	9/41 (55%)	11/37 (43%)
92-104	9/37 (47%)	16/31 (26%)	13/31 (33%)	6/26 (33%)
Term. Sac.	28	15	19	19

[1] number of deaths / number at risk

[2] Kaplan-Meier estimate of cumulative survival percentage (not the percentage corresponding to number of deaths / number at risk).

For display, survival in the two control groups is shown separately. However, the tests of homogeneity in survival based on the pooled control groups. Note there seems to be no statistically significant evidence of differences in survival across treatment groups.

The similar table (Table 5) for females is given below:

Table 5. Summary of Female Mice Mortality: Number/At Risk (KM estimate)

Period (Weeks)	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
1-50	2/60 (97%)	4/60 (93%)	2/60 (97%)	1/60 (98%)
51-78	10/57 (80%)	15/55 (68%)	12/56 (76%)	22/59 (62%)
79-91	9/47 (64%)	9/41 (53%)	11/44 (57%)	14/37 (38%)
92-104	9/38 (49%)	11/32 (35%)	10/32 (39%)	10/23 (22%)
Term. Sac.	29	21	22	13

[1] number of deaths / number at risk

[2] Kaplan-Meier estimate of cumulative survival percentage (not the percentage corresponding to number of deaths / number at risk).

Note that generally the control has the highest survival, while the low and medium dose groups track quite closely. The high dose groups have the highest mortality. These differences are quite statistically significant in females, and close to significance in males.

Tumorigenicity analysis:

Prior to adjusting for multiplicity the tests of trend in bronchio-alveola adenoma in the lungs male mice and benign skin pilomatricoma in female mice were statistically significant ($p \leq 0.0344$ and $p \leq 0.0345$, respectively). As discussed in section 1.1.3, there is a question about the appropriateness of these tests. However, assuming they are appropriate the Haseman-Lin-Rahman rules indicate that to adjust for multiplicity, for an overall roughly 10% Type I error rate, trends in common tumors should be tested at a 0.005 level and rare tumors at a 0.025 level. Using the incidence in the control group as a guide, bronchio-alveola adenoma in the lungs of male mice would be considered common while benign skin pilomatricoma in female mice would be considered as a rare tumor. Thus neither trend would be classified as statistically significant.

3.2.2. Study [redacted]-382009: A 24-Month Oral (Gavage) Carcinogenicity Study of MiraLax™ PEG-3350 in Rats.

As in the mouse study, there were four treatment groups with initially the same doseages as in the mouse study (i.e., a control, and three nominal dosages of MiraLax: Control, 1.5, 3, and 6 g/kg/day), labeled as Control, Low, Medium, and High dose groups respectively. Each group initially had 65 animals per gender. Up to 15 animals were chose for euthanasia in an interim sacrifice at week 26. The Sponsor notes that "Dose administration in the 3 and 6 g/kg/day female groups was discontinued during study week 98, when the survival reached the protocol-specified survival of 20 animals within any group, All the surviving females in each group [were] then euthanized during study week 103, when the protocol-specified survival of 15 animals was reached in the 3 g/kg/day female group. The discontinuation of dosing during study week 98 for 3 and 6 g/kg/day female groups or the early termination during study week 103 for all surviving females was not attributed to the test article as survival was similar in the control group." (page 14 of Sponsor's report)

Animals were housed individually with food and water available ad libitum. The Sponsor reports that weights were recorded biweekly. However, the summary weight table provided in the Sponsor's report starts at Week 26. The Sponsor reports that microscopic examination was performed on all animals euthanized in extremis or found dead, as well as all animals in the Control and High dose groups. The study was initiated on July1, 2002 at a laboratory in [redacted] with dosing starting on July 9, 2002. The study was formally completed on April 20, 2005.

3.2.2.1 Sponsor's Results and Conclusions for Rats

This section presents a summary of the Sponsor's analysis of survivability and tumorigencity in rats.

Survival analysis:

The following table (Table 6) displays, for each treatment group, the number of animals eho died of natural causes, the number of animals at risk. And the Kaplan-Meier estimate of the proportion who survived at the end of the specified period. The p-value under the column labeled control is a dose response trend test. The other p-values correspond to pairwise tests with control. Note that neither descriptively, nor from tests of homogeneity, is there any particular evidence of treatment differences in mortality.

Table 6. Summary of Mortality in Rats: Deaths/At Risk (KM)

Males Time Interval	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
0-50	3/65 (95%)	6/65 (90%)	4/65 (93%)	2/65 (97%)
Interim Sacrifice	12	12	12	12
51-80	9/49 (78%)	9/47 (78%)	6/48 (82%)	7/49 (83%)
81-EOS	20/40 (39%)	19/38 (36%)	18/42 (47%)	17/42 (49%)
Terminal Sacrifice	20	19	24	25
p-value	Trend: 0.2192	vs Ctrl: 0. 6527	vs Ctrl: 0.5248	vs Ctrl: 0.4224

Females Time Interval	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
0-50	0/65 (100%)	3/65 (94%)	3/65 (94%)	1/65 (98%)
Interim Sacrifice	15	14	13	10
51-80	10/50 (80%)	14/47 (66%)	12/47 (69%)	15/50 (69%)
81-EOS	18/40 (44%)	13/32 (39%)	18/33 (32%)	17/35 (35%)
Terminal Sacrifice	22	19	15	18
p-value	Trend: 0.2578	vs Ctrl: 0. 3312	vs Ctrl: 0.1194	vs Ctrl: 0.1724

Tumorigenicity analysis:

The Sponsor also conducted Peto type analyses of tumorigenicity in rats (see Tables A.3.3 and A.3.4 in Appendix 2). As with mice, even prior to adjusting for multiplicity, there were no statistically significant differences in the incidence of neoplasms.

3.2.2.2 FDA Reviewer's Results

This section summarizes the Agency results on survival and tumorigenicity in male and female rats.

Survival analysis:

Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1. From these survival curve plots, the high dose group generally has higher mortality, particularly among males. However no treatment differences were statistically significant (all $p>0.500$, except for the comparison between the low dose group and the pooled controls, $p = 0.302$).

These results are summarized in the following tables (Tables 7 and 8). The data are grouped for the specified time period, and give the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage is the percent surviving at the end of the interval, as estimated using a Kaplan-Meier estimate on the ungrouped data.

Table 7. Summary of Male Rat Mortality: Number/At Risk (KM estimate)

Period (Weeks)	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
1-50	3/65 (95%)	6/65 (92%)	4/65 (93%)	2/65 (97%)
Interim Sacrifice	12	12	12	12
51-78	9/49 (78%)	8/47 (74%)	5/48 (84%)	7/49 (83%)
79-91	5/40 (68%)	5/39 (65%)	11/43 (62%)	8/42 (67%)
92-104	15/35 (39%)	25/34 (36%)	8/32 (47%)	9/34 (49%)
Terminal Sacrifice	20	19	24	25

[1] number of deaths / number at risk
[2] Kaplan-Meier estimate of cumulative survival percentage (not the percentage corresponding to number of deaths / number at risk).

Note that neither the logrank nor the Wilcoxon tests of homogeneity in survival were statistically significant (p = 0.5134 and p = 0.5468, respectively), quite consistent with notion of homogeneity in survival..

The similar table for females is given below in Table 8:

Table 8. Summary of Female Rat Mortality: Number/At Risk (KM estimate)

Period (Weeks)	Control	Low 1.5 g/kg/day	Medium 3 g/kg/day	High 6 g/kg/day
1-50	0/65 (100%)	3/65 (94%)	3/65 (94%)	1/65 (98%)
Interim Sacrifice	15	14	15	10
51-78	9/50 (82%)	13/47 (68%)	10/47 (74%)	14/50 (71%)
79-91	10/41 (62%)	7/33 (39%)	9/35 (55%)	10/36 (51%)
92-104	9/31 (44%)	7/26 (39%)	11/26 (38%)	10/26(35%)
Terminal Sacrifice	22	19	15	18

[1] number of deaths / number at risk
[2] Kaplan-Meier estimate of cumulative survival percentage (not the percentage corresponding to number of deaths / number at risk).

Again, neither the logrank nor the Wilcoxon tests of homogeneity in survival were statistically significant (p = 0.5495 and p = 0.4559, respectively), quite consistent with notion of homogeneity in survival..

As noted before, strictly speaking, lack of evidence of heterogeneity in survival should not be treated as proof of homogeneity in survival. Nontheless, it does seem indicative of homogeneity in survival in both rat genders.

Tumorigenicity analysis:

In rats, also prior to adjusting for multiplicity, the tests of trend in benign islet cell carcinoma in the pancreas of males and undifferentiated malignant carcinoma in the mammary gland of females were statistically significant or close to it ($p \leq 0.0432$ and $p \leq 0.0517$, respectively). However as discussed in Section 1.3.1, (1) the assumptions of the tests of trend are not clearly satisfied so the reported p-values may not be appropriate, and (2) even if appropriate, upon adjusting for multiplicity using the Haseman-Lin-Rahman rules above, none would be considered statistically significant. So no tests of tumorigenicity would be considered to be statistically significant. Again, absence of proof is not proof of absence. Nonetheless these results are consistent with the notion of no particular carcinogenic signal.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5 SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

5.1.1 Statistical Issues

Please see Section 1.3 above.

5.1.2 Collective Evidence

In both genders in mice the high dose group generally had the highest mortality rate, while the low and medium dose groups were more or less similar in mortality, but with somewhat higher mortality than in the control group. In rats there were no clear differences in survival among the treatment groups. The significance levels of the tests of homogeneity among the treatment groups are presented in Table 9 below (please see Appendix 1 for details).

Table 9. (identical to Table 2) Tests of Homogeneity and Trend in Survival

	Mice			Rats		
Gender	Log Rank	Wilcoxon	Trend	Log Rank	Wilcoxon	Trend
Male	0.0826	0.0506	0.1745	0.5134	0.5648	0.1673
Female	0.0062	0.0064	0.0064	0.5495	0.4559	0.3047

Note that in mice the tests of homogeneity in survival were clearly statistically significant in females and close to significance in male mice. However, there were no treatment group related differences in mortality that were apparent in rats.

For the tests for tumorigenicity, in both species, the only statistically significant results were in trend tests. Even without adjusting for multiplicity, there were no statistically significant pairwise differences between the control group and the high dose group. In mice, prior to adjusting for multiplicity, the tests of trend in bronchio-alveola adenoma in the lungs males and benign skin pilomatricoma in females were statistically significant ($p \leq 0.0344$ and $p \leq 0.0345$, respectively). In rats, also prior to adjusting for multiplicity, the tests of trend in benign islet cell carcinoma in the pancreas of males and undifferentiated malignant carcinoma in the mammary gland of females were statistically significant or close to it ($p \leq 0.0432$ and $p \leq 0.0517$, respectively). However as discussed in Section 1.3.1 above, (1) the assumptions of the tests of trend are not clearly satisfied and thus the reported p-values may not be appropriate, and (2) even if appropriate, upon adjusting for multiplicity using the Haseman-Lin-Rahman rules, none would be considered statistically significant. So no tests of tumorigenicity would be considered to be statistically significant.

5.2. Conclusions and Recommendations

In both studies there were four treatment groups (i.e., a control, and three nominal dosages of MiraLax: Control, 1.5, 3, and 6 g/kg/day), labeled as Control, Low, Medium, and High dose groups respectively. In both studies treatment in the high dose groups in females was terminated early.

For both genders in mice, the high dose group generally had the highest mortality rate, while the low and medium dose groups were more or less similar in mortality, but with somewhat higher mortality than in the control group. These differces were statistically significant in female mice, and close to significant in males. There were no clear differences in mortality among the rat treatment groups in either gender.

For the tests for tumorigenicity, in both species, the only statistically significant results were in trend tests. The trend tests are biased due to the fact that not all animals were analyzed. However, even if the trend tests are accepted, after adjusting for multiplicity there were no statistically significant pairwise differences between the control group and the high dose group and no statistically significant tests of trend. Absence of proof is not proof of absence. Nonetheless these results are consistent with the notion of no particular carcinogenic signal.

The combination of the body weight gain data and the mortality information in the studies indicate that the high dose used in the mouse and the female rat studies are close to the maximum Tolerated Dose (MTD). The evaluation of the appropriateness of the high doae in the female rat study is inconclusive. Information regarding clinical signs and histopathological data, plus other possible considerations, are well beyond the expertise of this reviewer, but presumably would be used by the toxicologist in the final assessment of the adequacy of these experiments.

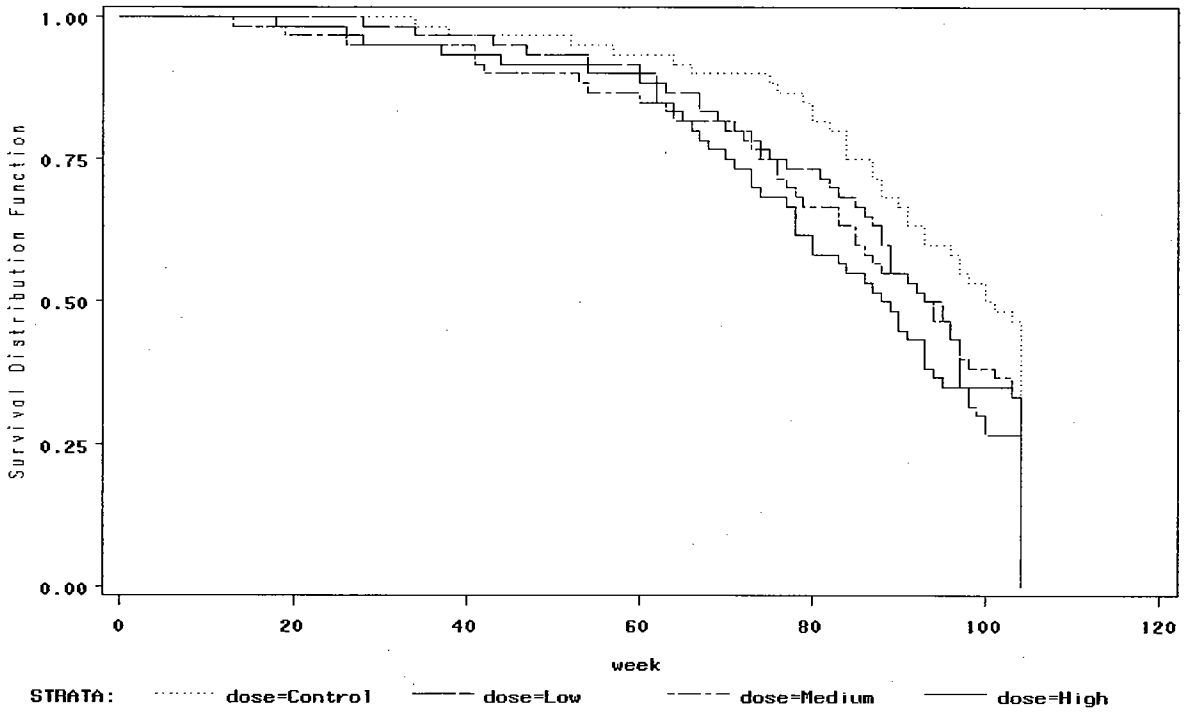
APPENDICES:

Appendix 1. Survival Analysis

In both genders in mice the high dose group generally had the highest mortality rate, while the low and medium dose groups were more or less similar in mortality, but with somewhat higher mortality than the control group. However in mice the tests of homogeneity in survival were only clearly statistically significant in females, consistent hypothesis of differences in survival (Females: Logrank p = 0.0062, Wilcoxon p = 0.0064, proportional hazards test of trend p = 0.0064). Differences were close to significance in male mice (Males: Logrank p = 0.0826, Wilcoxon p = 0.0506, trend p = 0.1745). Treatment group related differences in mortality were not apparent in rats (Males: Logrank p = 0.5134, Wilcoxon p = 0.5648, proportional hazards test of trend p = 0.1673, Female: Logrank p = 0.5495, Wilcoxon p = 0.4559, proportional hazards test of trend p = 0.3047).

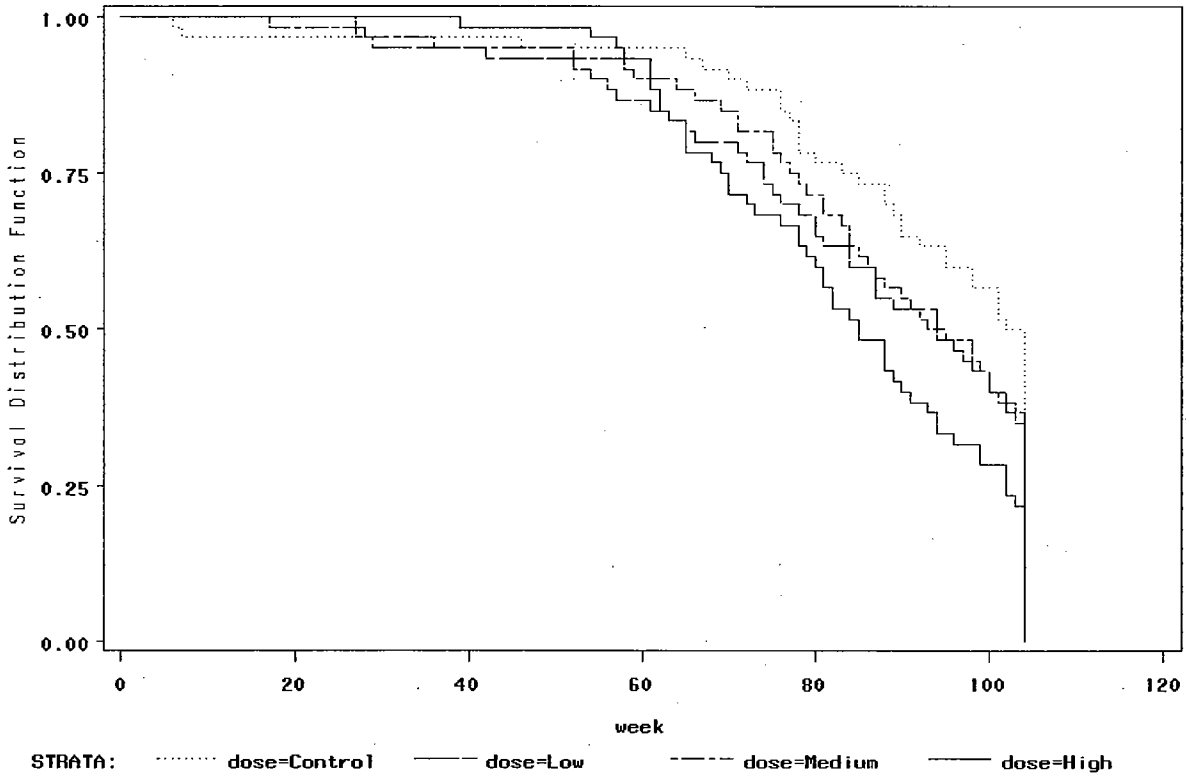
The figures below display the Kaplan-Meier estimated survival curves for the four different species by gender combinations. These curves include the time of censoring, including sacrifice or acidental death, as an event.

Figure A.1.1 Male Mice



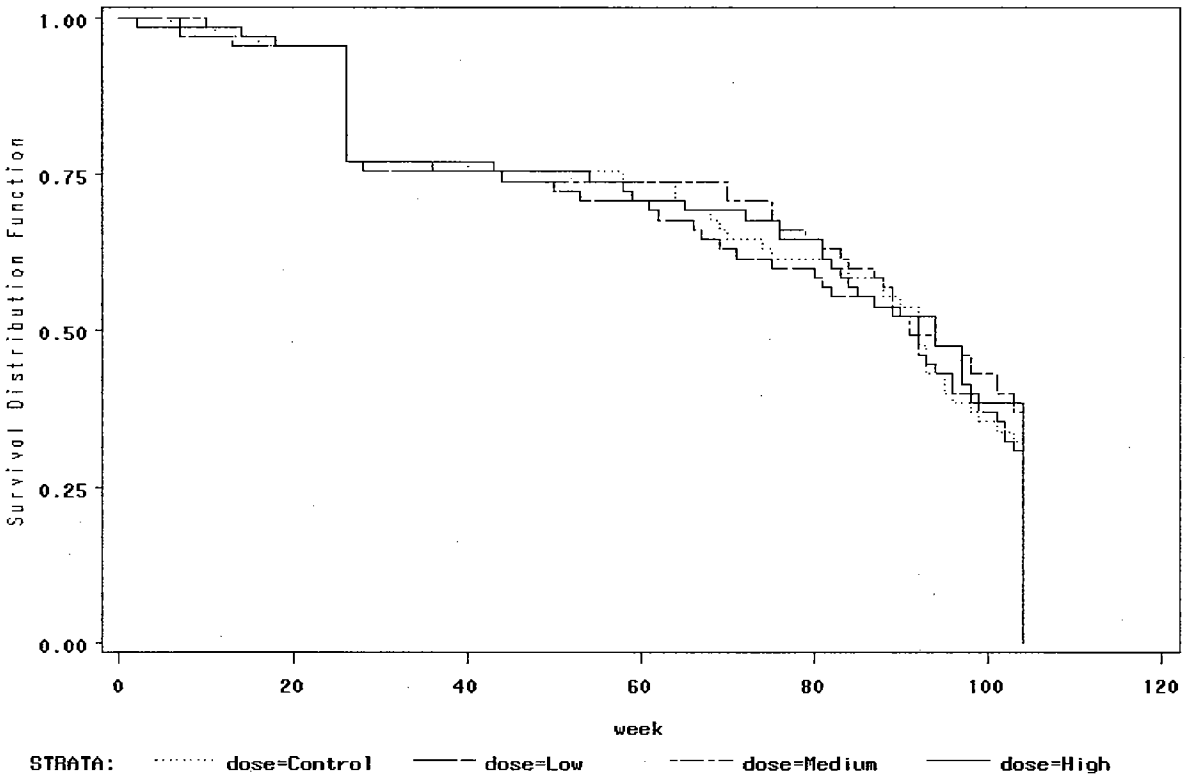
Note above that the tests of homogeneity in survival in male mice were close to being statistically significant (Logrank p = 0.0826, Wilcoxon p = 0.0506, and Cox: p = 0.1742). Recall, however, that the tests of significance treat censored data as reductions in the risk set, not as events. However they are treated as events in the survival plots presented here. It was felt that, while not appropriate for testing, this gave a better picture of the survival over the course of the study.

Figure A.1.2 Female Mice



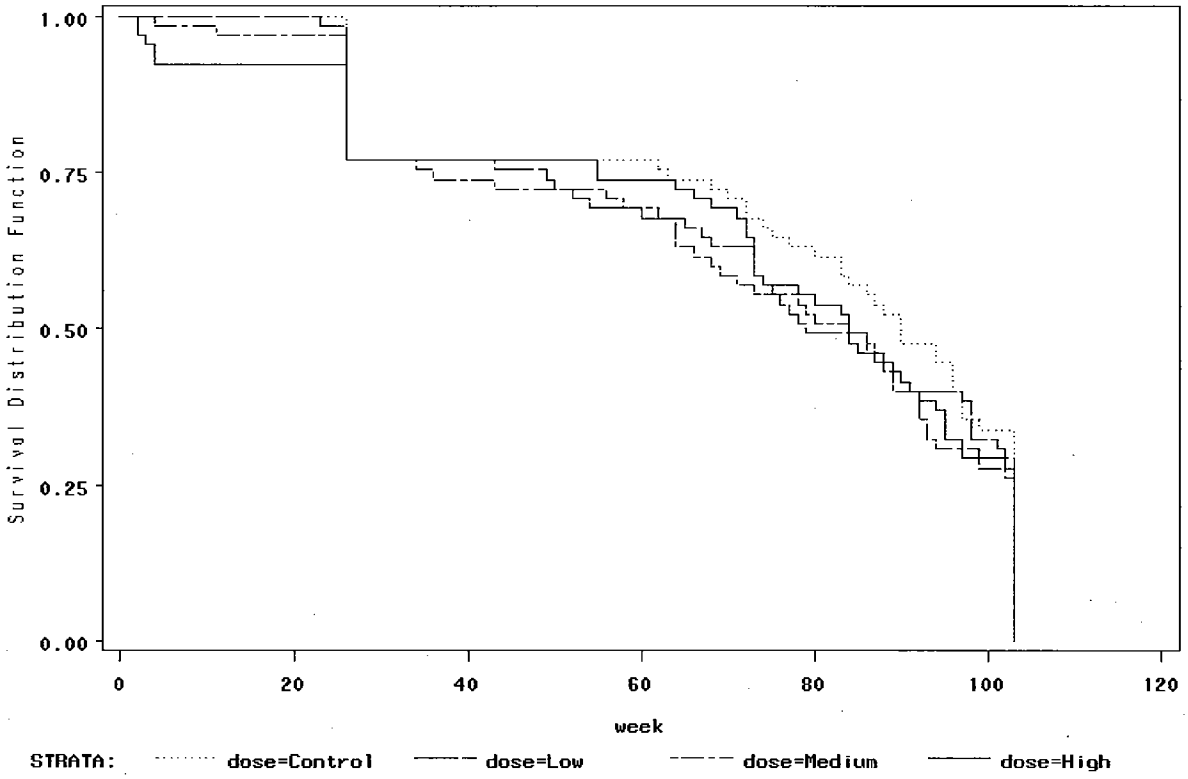
In female mice, tests of homogeneity in survival were statistically significant (Logrank p = 0.0062, Wilcoxon p = 0.0064, and Cox: p = 0.0064).

Figure A.1.3 Male Rats



As noted above, in male rats, tests of homogeneity in survival were clearly not statistically significant (Logrank p = 0.5134, Wilcoxon p = 0.5648, and Cox: p = 0.1673).

Figure A.1.4 Female Rats



In female rats, tests of homogeneity in survival were also clearly not statistically significant (Logrank p = 0.5495, Wilcoxon p = 0.4559, and Cox: p = 0.3047).

Appendix 2. Bayesian Analysis of Survival

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

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

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

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

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:

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

Note that without other information we would expect the treatment effects in the control groups to be exchangeable (i.e., effectively the treatment groups can be treated as identical). That is, treatment effects are investigated by assessing the effect differential effects of treatment in the $exp(x'\beta)$ term.

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. That is, suppose the time axis can be partitioned as $(a_1=0, a_2]$, $(a_2, a_3]$, . . ,$(a_T,a_{T+1}]$. Assume a constant baseline hazard λ_j for observations in $(a_j,a_{j+1}]$.

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). Thus, for identification, in the mouse study there are only four degrees of freedom for testing mortality differences among the five treatment groups. In the rat study there are three degrees of freedom for testing mortality differences among the four treatment groups. If we confound specification of the baseline hazard with the pooled controls, then treatment effects over the remaining treatments correspond to differences from controls. Further, using the trend specification, we can confound the baseline hazard with the intercept, which in turn defines the effect of the control (i.e., dose=0). This can be expressed mathematically as follows:

Using so called dummy coding, we can define, for each treatment group k,

$\delta_k = 1$ for the ith treatment group,

0 otherwise.

Then three possibly relevant models for treatment effect could be expressed as follows:

(1) Parameterization of a different effect for each treatment (with 5 treatments),

$x_i'\beta = \beta_0 + \beta_1 * \delta_1 + \beta_2 * \delta_2 + \beta_3 * \delta_3 + \beta_4 * \delta_4$.

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

$x_i^t\beta = \beta_0 + \beta_1 *$ dose

(3) Parameterization of no differences in survival across treatment groups, $x_i^t\beta = \beta_0$.

Note that for each of these models $\exp(\beta_0)$ is confounded with the baseline hazard $h_0(t)$ and is not estimated. In the programs below, the other β_k is denoted by beta[k] (or beta when only the slope term is used). In model 1), with this coding, the effect of the difference between treatment i and the pooled controls is assessed by the β_k.

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

$$H_o(t_i) = e^{x'\beta}\int_0^{t_i} h_0(u)du = e^{x'\beta}\{\sum_{k=1}^{j-1}\lambda_k(a_k - a_{k-1}) + \lambda_j(t_i - a_{j-1})\},$$

with hazard $h_o(t_i) = e^{x'\beta}\lambda_j$.

Then the likelihood for subject i can be written as:

$$L_i(\lambda,\beta) \propto \begin{cases} e^{-H_0(t_i)} & \textit{if ith subject is censored at time } t_i \\ \lambda_j e^{x'\beta} e^{-H_0(t_i)} & \textit{if ith subject fails at time } t_i \end{cases}$$

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

For subject i censored or failed at time t_j, let $\gamma_{ik} = \begin{cases} \lambda_k(a_k - a_{k-1}) & \textit{for } t_j > a_k \\ \lambda_j(t_j - a_{j-1}) & \textit{for } a_{j-1} \le t_j < a_j \\ 0 & \textit{otherwise} \end{cases}$

Note since the subject i is censored or failed at time t_j, for intervals above a_j, $-e^{x'\beta}\gamma_{ik} = 0$.

Then for these intervals, $\exp(-e^{x'\beta}\gamma_{ik})$ does not contribute to the product.

Thus $S(t) = e^{-H(t)} = \prod_{k=1}^{T}\exp(-e^{x'\beta}\gamma_{ik})$. Further, with respect to parameters $(t_j - a_{j-1})$ is constant, and hence can be incorporated in the likelihood for subjects who fail by multiplying λ_j by this difference. Thus, for subject i, the likelihood can also be written as:

$$L_i(\lambda,\beta) \propto \begin{cases} \prod_{k=1}^{T} \exp(-e^{x'\beta}\gamma_{ik}) & \text{if ith subject is censored at time } t_i \\ \gamma_{ij} e^{x'\beta} \prod_{k=1}^{T} \exp(-e^{x'\beta}\gamma_{ik}) & \text{if ith subject fails at time } t_i \end{cases}$$

Note this corresponds to the likelihood of T independent Poisson random variables with mean $e^{x'\beta}\gamma_{ik}$ where all responses are zero except at time j with the occurrence of a failure in the jth interval $(a_{j-1},a_j]$. This is only a computational convenience but allows easy estimation of the appropriate parameters using standard software (e.g., WINBUGS)..

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. A gamma prior would be skewed to the right and would seem to be an appropriate choice. The two year study is broken down into thirteen 56 day periods. Sacrifice or accidental death is treated as a reduction in the risk set, but not a mortality event. For convenience we specify the same prior over each period. We might expect mortality to occur in a little over half the subjects. So it would seem to be reasonable to specify an expected baseline hazard for a subject to be about 0.05 in each of the 13 periods. However, the chance of an event can increase considerably, so we specify a variance of about .25. Under the parameterization used by WINBUGS this corresponds to a Gamma(0.01, 0.2) distribution, as is used in the programs below.

One possible approach for model selection in Bayesian models is to use the Deviance Information Criterion (DIC). Effectively, for $D(\theta)$ denoting the usual deviance, $DIC \approx E(D(\theta)) + 1/2\,(Var\,(D(\theta)))$. 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 (from WINBUGS) are given below:

Deviance Information Criterion for Mice	Males	Females
Model with heterogeneity over the four treatment groups	15.428	14.510
Model with linear trend in dose groups, 0=control	13.456	12.341
Model with constant dose effect	12.408	11.382

Deviance Information Criterion for Rats	Males	Females
Model with heterogeneity over the four treatment groups	16.560	14.596
Model with linear trend in dose groups, 0=control.	14.626	12.599
Model with constant dose effect	13.623	11.453

Using the DIC, for both genders and both species species, and for male mice, the models with no treatment effect on survival seems to fit best.

However, the tables below summarize the estimated posterior distributions of the treatment parameters. For male mice the approximate credible intervals (i.e., with lower endpoint in the 2.5% column and upper endpoint in the 97.5% column) do not include 0 for the parameters denoting the difference between the high dose and low dose groups and the control. That should be considered as reasonably strong evidence of treatment differences. For female mice the posterior distribution of the difference in treatment effect between the high dose group and the control is even more concentrated away from zero. Further, for females the posterior distribution of the parameter denoting trend is concentrated away from zero, suggesting a "significant" treatment effect. It is not clear if this is a weakness in the DIC, but is an issue that needs research.

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

Male testing homogeneity over five parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.5328	0.2374	0.005877	0.07414	0.5303	0.9965	4001	12000
beta[2]	0.4088	0.244	0.006146	-0.0726	0.4086	0.895	4001	12000
beta[3]	0.5224	0.2443	0.005902	0.03972	0.5227	0.9978	4001	12000

Male model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	0.06331	0.03549	6.902E-4	-0.007192	0.06361	0.133	4001	12000

Female testing homogeneity over five parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.451	0.2495	0.00553	-0.03238	0.4518	0.9433	4001	12000
beta[2]	0.2989	0.2554	0.005346	-0.199	0.2982	0.8053	4001	12000
beta[3]	0.785	0.2419	0.005095	0.311	0.7844	1.264	4001	12000

Female model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	0.1147	0.03587	6.391E-4	0.04421	0.1147	0.1855	4001	12000

For rats the posterior summaries of the parameter values are given in Table A.2.2 below. Note that the posterior distributions of the parameters are not concentrated away from zero. That is, in each case the approximate 95% credible interval for the parameter includes zero. This should be considered as evidence that treatment does not have a strong effect on survival.

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

Male testing homogeneity over four parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.07789	0.2461	0.004483	-0.4036	0.08093	0.5541	4001	12000
beta[2]	-0.2038	0.2595	0.004617	-0.7089	-0.2068	0.3098	4001	12000
beta[3]	-0.2568	0.2668	0.004806	-0.7864	-0.2561	0.2582	4001	12000

Male model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	-0.05296	0.04234	7.42E-4	-0.1367	-0.0524	0.0293	4001	12000

Female testing homogeneity over four parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.1957	0.2685	0.006248	-0.3181	0.1925	0.7256	4001	12000
beta[2]	0.3485	0.2622	0.006373	-0.1592	0.344	0.8639	4001	12000
beta[3]	0.2906	0.2626	0.006274	-0.2215	0.2875	0.8177	4001	12000

Female model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	0.04315	0.03872	6.979E-4	-0.03481	0.04366	0.1173	4001	12000

Thus, despit the indications of the DIC there seems to be evidence of treatment effects on mortality in mice., particularly female mice, but no particular evidence in rats of either gender.

Due to severe time constraints there was no detailed, systematic attempt to assess convergence of the MCMC iterations or to assess model fit. However, the autocorrelations were quite low, the history plots showed good mixing, and the posterior distributions were approximately symmetric and seemed to follow normal distributions. So, given the model, these should be reasonable estimates.

Programs similar to the following were used in the analyses:

Testing homogeneity over four parameter groups (for Mice):

```
model{
    for (j in 1:T+1) {
          a[j] <-  ( j-1)*56+1
                   }
   for (i in 1:N) {
       lin.pred[i]  <-  beta[1]*equals(dose[i],2)+ beta[2]*equals(dose[i],3)+
                        beta[3]*equals(dose[i],4)
   for (j in 1:T) {
          d[i,j]<- fail[i]*step(obs.t[i]-a[j])*step(a[j+1]-obs.t[i])
          gamma[i,j] <- (a[j+1]-a[j])*step(obs.t[i]-a[j+1])+(obs.t[i]- a[j])
                  *step(a[j+1]-obs.t[i])*step(obs.t[i]-a[j])
          theta[i,j] <-  lambda[j] * exp(lin.pred[i])
          d[i,j]~ dpois(mu[i,j])
          mu[i,j] <- theta[i,j]*gamma[i,j]
                   }
                 }
   for ( j in 1:T) {
           lambda[j] ~ dgamma(0.01,0.2)
            part[j] <- lambda[j]*(a[j+1]-a[j])
                      }
   for (m in 1:4)  {
           beta[m] ~ dnorm (0.0 ,  0.001)
                          }
   for ( k in 1:T) {
           sum[k] <- sum(part[1:k])
           S.high[k] <- exp( -(exp(beta[3])*sum[k]))
           S.med[k]  <- exp( -(exp(beta[2])*sum[k]))
           S.low[k]  <- exp( -(exp(beta[1])*sum[k]))
           S.veh[k]  <- exp( -(sum[k]))
                    }
   }
inits
list(beta=c(-1,0,1))

data
list(N=240,T=13)
dose[ ] obs.t[ ] fail[ ]
 1     48   1
 1    728   0
 1    686   1
   - data -
 4   728   0
END
```

Assessing simple trend:

```
model{
    for (j in 1:T+1) {
          a[j] <-  ( j-1)*56+1
                   }
   for (i in 1:N) {
       lin.pred[i]<- beta*(1.5*equals(dose[i],2)+ 3*equals(dose[i],3)+
                  6*equals(dose[i],4))
   for (j in 1:T) {
          d[i,j]<- fail[i]*step(obs.t[i]-a[j])*step(a[j+1]-obs.t[i])
          gamma[i,j] <- (a[j+1]-a[j])*step(obs.t[i]-a[j+1])+(obs.t[i]-a[j])
                     *step(a[j+1]-obs.t[i])*step(obs.t[i]-a[j])
          theta[i,j] <-  lambda[j] * exp(lin.pred[i])
          d[i,j]~ dpois(mu[i,j])
          mu[i,j] <- theta[i,j]*gamma[i,j]
                   }
                 }
    for ( j in 1:T) {
            lambda[j] ~ dgamma(0.01,0.2)
            part[j] <- lambda[j]*(a[j+1]-a[j])
                    }
    beta ~ dnorm (0.0 ,  0.001)

    for ( k in 1:T) {
           sum[k] <- sum(part[1:k])
           S.high[k]   <- exp( -(exp(6*beta)*sum[k]))
           S.med[k]    <- exp( -(exp(3*beta)*sum[k]))
           S.low[k]    <- exp( -(exp(1.5*beta)*sum[k]))
           S.veh[k]    <- exp( -(sum[k]))
                  }
   }

inits
list(beta=0.01)

data
list(N=240,T=13)
dose[ ] obs.t[ ] fail[ ]
 1     48   1
 1    728   0
 1    686   1
   - data -
 4   728   0
END
```

Appendix 3. Sponsor's Tumorigenicity Analyses

For each gender within each species, for each neoplasm within each organ, the Sponsor provided tables of tumor incidence and the results of an analysis of tumor incidence using Peto's mortality-prevalence method, stratifying on the context in which animals were observed. The Sponsor reports that for organs in which an exhaustive examination of animals was planned the results of one-sided tests of trend are presented. Otherwise, the results of a pairwise test between the controls and the high dose group are presented.

The Haseman-Lin-Rahman rules were used to adjust for the multiplicity of comparisons. 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 each type of comparison, one compares the unadjusted significance level to the appropriate bound below:

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

The Sponsor indicated that the historical incidence rate was used to determine if a tumor was rare or not. However, even before adjusting for multiplicity, in both studies no tests of differences or tests of trend were statistically significant.

There were differences between the exact p-values computed in the FDA analysis and those provided by the Sponsor. However, after adjusting for multiplicity, either by the Sponsor's methods or by the Agency's, these differences do not seem to have an impact on conclusions.

Table A.3.1. [redacted] -382018 Tumor Incidence and Tests of Trend/Incidence in Male Mice

Males	Control	Low	Medium	High	P-value
Adrenal Cortex	60			58	
Cortex, Adenoma	1	-	-	1	0.6348
Adrenal Medulla	57			54	
Pheochromocytoma, Benign	1	-	-	0	1.00
Epididymides	60			60	
Granular Cell Tumor, Benign	0	-	-	1	0.7826
Harderian Glands	59			60	
Adenoma	10	-	-	7	0.6601
Carcinoma	1	-	-	0	1.00
Adenoma/Carcinoma	11	-	-	7	0.7344
Kidneys	60	60	60	59	
Adenoma, Renal Tubule	0	2	1	1	0.3198
Carcinoma, Renal Tubule	1	0	0	0	1.000
Adenoma/Carcinoma, Renal Tubule	1	2	1	1	0.5219

Liver	60			60	
Adenoma, Hepatocellular	15	-	-	6	0.9107
Carcinoma, Hepatocellular	4	-	-	4	0.4246
Adenoma/Carcinoma, Hepatocellular	19	-	-	10	0.7793
Lungs	60			60	
Adenoma, Bronchio Alveolar	8	-	-	12	0.0986
Carcinoma, Bronchio Alveolar	4	-	-	4	0.4545
Adenoma/Carcinoma, Bronchio Alveolar	11	-	-	15	0.0606
Pancreas	60			58	
Adenoma, Islet Cell	1	-	-	0	1.00
Pituitary	56			57	
Carcinoma, Pars Distalis	0	-	-	1	0.4318
Skin	60			59	
Papilloma, Squamous Cell	0	-	-	1	0.4118
Spleen	60			60	
Sarcoma, Undifferentiated	0	-	-	1	0.4231
Stomach, Glan	60			60	
Neuroendocrine Cell Tumor, Malignantl	0	-	-	1	0.4375
Systemic	60	60	60	60	
Sarcoma, Histiocytic	1	0	2	2	0.1571
Mesothelioma, Malignant	1	0	0	0	1.00
Lymphoma, Malignant	5	2	2	4	0.4792
Hemangiosarcoma	4	5	5	4	0.4047
Hemangioma	2	2	5	3	0.2370
Hemangiosarcoma/Hemangioma	6	7	10	7	0.2594
Thyroid Glands	60			59	
Adenoma, Follicular Cell	0	-	-	1	0.7826
Carcinoma, Follicular Cell	1	-	-	1	0.8677
Adenoma/Carcinoma, Follicular Cell	1	-	-	2	0.7439

Table A.3.2 — 382018 Tumor Incidence and Tests of Trend/Incidence in Female Mice

Males	Control	Low	Medium	High	P-value
Adrenal Cortex	60			60	
Adenoma, A Cell	1	-	-	1	0.5285
Adenoma	1	-	-	0	1.00
Adrenal Medulla	45			53	
Pheochromocytoma, Benign	0	-	-	1	0.6500
Cervix	60			55	
Leiomysarcoma	1	-	-	0	1.00
Carcinoma, Squamous Cell	1	-	-	0	1.00
Fibroma	1	-	-	0	1.00
Harderian Glands	60	60	60	60	
Adenoma	1	1	8	3	0.1064
Carcinoma	1	1	0	0	0.8863
Adenoma/Carcinoma	2	2	8	3	0.1954
Kidneys	60	60	60	60	
Renal Mesenchymal Tumor, Malignant	0	0	1	0	0.6087
Liver	60			60	
Adenoma, Hepatocellular	2	-	-	1	0.7766

Lungs	60			60	
Adenoma, Bronchio Alveolar	6	-	-	5	0.7600
Carcinoma, Bronchio Alveolar	6	-	-	6	0.4864
Adenoma/Carcinoma, Bronchio Alveolar	10	-	-	11	0.2864
Mammary Gland	54			55	
Adenocarcinoma	1	-	-	2	0.4853
Ovaries	54			60	
Luteoma, Benign	1	-	-	0	1.00
Granulosa Cell Tumor, Benign	1	-	-	0	1.00
Cystadenoma	0	-	-	2	0.2177
Adenoma, Tubulostromal	2	-	-	0	1.00
Granulosa Cell Tumor, Benign/Malignant	1	-	-	0	1.00
Pituitary	57			57	
Adenoma, Pars Distalis	1	-	-	0	1.00
Skin	60			59	
Sarcoma, Undifferentiatedl	0	-	-	1	0.4362
Pilomatricoma	0	-	-	2	0.1890
Stomach, Glan	60			60	
Polyp	1	-	-	0	1.00
Systemic	60	60	60	60	
Sarcoma, Histiocytic	5	9	9	4	0.5234
Lymphoma, Malignant	13	10	8	12	0.1864
Leukemia, Granulocytic	0	2	0	0	0.7691
Hemangiosarcoma	4	5	6	6	0.2272
Fibrous Histiocytoma, Malignant	0	0	0	1	0.1780
Hemangioma	9	5	3	6	0.9308
Hemangiosarcoma/Hemangioma	13	10	9	12	0.6958
Uterus	60			60	
Sarcoma, Endometrial Stromal	1	-	-	0	1.00
Leiomyosarcoma	3	-	-	0	1.00
Carcinoma	3	-	-	1	0.9361
Polyp, Endometrial Stromal	4	-	-	6	0.3565
Leiomyoma	1	-	-	1	0.8520
Polyp/Sarcoma, Endometrial Stromal	5	-	-	6	0.3213
Leiomyoma/ Leiomyosarcoma	4	-	-	1	0.9894
Uterus/Cervic	60			60	
Sarcoma, Endometrial Stromal	1	-	-	0	1.00
Leiomyosarcoma	4	-	-	0	1.00
Leiomyoma	1	-	-	1	0.8520
Polyp/Sarcoma, Endometrial Stromal	5	-	-	6	0.3213
Leiomyoma/ Leiomyosarcoma	5	-	-	1	0.9927

Tables A.3.3 and A.3.4 below summarize the tumor incidence and results of the tests in rats for each treatment group.

Table A.3.3 [redacted]-382009 Tumor Incidence and Tests of Trend/Incidence in Male Mice

Males	Control	Low	Medium	High	P-value
Adrenal Medulla	61			62	
Pheochromocytoma, Benign	2	-	-	5	0.1816
Pheochromocytoma, Malignant	0	-	-	3	0.1621

Pheochromocytoma, Benign/Malignant	2	-	-	6	0.1137
Brain	65			64	
Astrocytoma, Malignant	0	-	-	2	0.2439
Femur	65			65	
Osteosarcoma	0	-	-	1	0.5536
Kidneys	59	59	56	63	
Liposarcoma	0	1	0	0	0.7727
Liver	65			65	
Adenoma, Hepatocellular	0	-	-	1	0. 5556
Pancreas	61			64	
Adenoma, Islet Cell	4	-	-	7	0.2406
Carcinoma, Isley Cell	1	-	-	1	0.8192
Parathyroids	58			57	
Adenoma	1	-	-	0	1.00
Pituitary	65			62	
Adenomoa, Pars Distalis	27	-	-	22	0.8397
Rectum	60			59	
Leiomyosarcoma	1	-	-	0	1.00
Skin	65			65	
Papilloma, Squamous Cell	0	-	-	1	0.4948
Lipoma	2	-	-	0	1.00
Fibroma	1	-	-	1	0.7827
Basal Cell Tumor	1	-	-	2	0.5078
Stomach, Non	65			65	
Papilloma, Squamous Cell	1	-	-	0	1.00
Systemic	65	65	65	65	
Sarcoma, Histiocytic	1	3	2	2	0.4738
Lymphoma, Malignant	1	0	2	0	0.7243
Leukemia, Granulocytic	0	0	0	1	0.2297
Testes	65			65	
Adenoma, Interstitial Cell	2	-	-	2	0.7387
Thymus	63			64	
Fibrosarcoma	1	-	-	0	1.00
Thyroid Glands	60			59	
Adenoma, C-Cell	4	-	-	2	0.8900
Carcinoma, C-Cell	1	-	-	2	0.5849
Urinary Bladder	62			63	
Leiomyoma	1	-	-	0	1.00

Table A.3.4 — 382009 Tumor Incidence and Tests of Trend/Incidence in Female Mice

Males	Control	Low	Medium	High	P-value
Adrenal Cortex	64			64	
Adenoma	1	-	-	0	1.00
Adrenal Medulla	64			65	
Pheochromocytoma, Benign	0	-	-	1	0.4500
Pheochromocytoma, Malignant	1	-	-	0	1.00
Brain	65			65	
Astrocytoma, Malignant	0	-	-	1	0.4500
Heart	65			65	
Schwannoma, Endocaedial, Benign	1	-	-	0	1.00

Jejunum	62			57	
Leiomyoma	1	-	-	0	1.00
Kidneys	64			57	
Lipoma	0	-	-	1	0.4500
Mammary Gland	62			63	
Fibrosarcoma	2	-	-	0	1.00
Carcinoma, Undifferentiated	0	-	-	2	0.2212
Adenocarcinoma	12	-	-	12	0.4635
Fibroadenoma	18	-	-	17	0.5059
Adenoma	3	-	-	2	0.7727
Nasal Level III	65			64	
Carcinoma, Squamous Cell	1	-	-	0	1.00
Ovaries	65			63	
Luteoma	1	-	-	0	1.00
Pancreas	65			63	
Adenoma, Islet Cell	2	-	-	0	1.00
Pituitary	65			65	
Adenomoa, Pars Distalis	40	-	-	36	0.6843
Skin	64			63	
Schwannoma, Malignant	0	-	-	1	0.4462
Basal Cell Tumor, Malignant.	0	-	-	1	0.4545
Lipoma	0	-	-	1	0.4500
Systemic	65			65	
Sarcoma, Histiocytic	1	-	-	0	1.00
Lymphoma, Malignant	0	-	-	1	0.4528
Leukemia, Granulocytic	2	-	-	3	0.4577
Thymus	65			64	
Hibernoma, Malignant	0	-	-	1	0.4762
Thyroid Glands	63			59	
Adenoma, Follicular Cell	2	-	-	0	1.00
Adenoma, C-Cell	2	-	-	4	0.2270
Uterus	64			64	
Schwannoma, Malignant	1	-	-	0	1.00
Carcinoma	0	-	-	1	0.4545
Polyp, Endometrial Stromal	3	-	-	4	0.4073
Vagina	64			63	
Leiomyoma	0	-	-	1	0.4545

Appendix 4. FDA Tumorigenicity Analysis

Tables A.4.1 and A.4.2 below display the number of neoplasms in each organ and tumor combination in mice taken from the datasets provided by the Sponsor. Tables A.4.3 and A.4.4 below display similar results for rats. For each dose group, the numbers in the table are the number of animals where histopathological analysis detected a tumor. For mice, the Sponsor notes that for most organs all 60 animals at risk in the dose group were only analyzed in the high dose group (6 g/kg/day) and the control group. For the low (1.5 g/kg/day) and medium (3 g/kg/day) dose groups only 40-50 selected animals were analyzed. As discussed in section 1.3.1 unless all animals at risk were examined or animals to be examined were chosen completely at random it is not clear how to adjust the analysis for the actual number of animals analyzed in the treatment group. The software used for the FDA analysis assumes that all organs were analyzed in all animals. Thus, the tests of significance for the tests of trend are computed under the assumption that those animals which were not analyzed did not manifest tumors. In the table below, for those organ-noplasm combinations where not all animals were analyzed the significance level of the test of no trend are presented, but in enclosed in parentheses. This is meant to indicate that, strictly speaking, the tests do not satisfy their assumptions, and hence are not actually appropriate. However, the number of animals analyzed in the low and medium dose groups is large enough that results might be considered suggestive, particularly statistically significant results. For those few organ-tumor combinations where all animals were analyzed the significance levels of the tests of trend can be expected to satisfy the assumptions of the test and are not enclosed in parentheses. The pairwise comparisons between the high dose group and control do not have this limitation and are not enclosed in parantheses. Finally, note that 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).

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 each type of comparison, one compares the unadjusted significance level to the appropriate bound below:

Haseman - Lin - Rahman Bounds: Comparison	Rare Tumor (Incidence ≤ 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 ≤ 1%, i.e. 0 or 1 tumor), a trend would be considered statistically significant if the computed significance level was at or less than 0.025, while a comparison between the high dose group and the pooled controls (i.e., a pairwise comparison) would be statistically significant if the computed significance level was no more than 0.05.

For both species, the only statistically significant results were in trend tests. Even without adjusting for multiplicity, there were no statistically significant pairwise differences between the control group and the high dose group. In mice, prior to adjusting for multiplicity the tests of trend in bronchio-alveola adenoma in the lungs males and benign skin pilomatricoma in females were statistically significant ($p \leq 0.0280$ and $p \leq 0.0345$, respectively). In rats, also prior to adjusting for multiplicity, the tests of trend in benign islet cell carcinoma in the pancreas of males and undifferentiates malignant carcinoma in the mammary gland of females were statistically significant or close to it ($p \leq 0.0432$ and $p \leq 0.0517$, respectively). However, 1) the assumptions of the tests of trend are not clearly satisfied so the reported p-values may not be appropriate, and 2) even if appropriate, upon adjusting for multiplicity using the Haseman-Lin-Rahman rules above, none would be considered statistically significant.

The following tables show the tumor incidence and the significance levels of the tests of trend and the high dose group versus the pooled controls. When there are no observed values in the controls and the high dose group, the test of differences is not defined and thus no p-value is given. The character string "#B" seems to denote benign tumors while "#M" denotes malignant tumors.

Table A.4.1. Tumorgenicity in Male Mice

Organ / Tumor	Control	Low	Medium	High	p-values: Trend	Hi vs Cntrl
ADRENAL CORTEX						
#B ADENOMA, A CELL	1	0	0	1	(0.4634)	0.6503
ADRENAL MEDULLA						
#B PHEOCHROMOCYTOMA, BENIGN	1	0	0	0	(1.0000)	1.0000
#M PHEOCHROMOCYTOMA, MALIGNANT	0	0	1	0	(0.4699)	
BONE						
#M OSTEOSARCOMA	0	0	1	0	(0.4762)	
EPIDIDYMIDES						
#B GRANULAR CELL TUMOR, BENIGN	0	0	0	1	(0.3673)	0.7500
HARDERIAN GLANDS						
#B ADENOMA	8	1	3	6	(0.4701)	0.6800
#B ADENOMA, MULTIPLE	2	0	0	1	(0.7672)	0.9162
#M CARCINOMA	1	0	0	0	(1.0000)	1.0000
KIDNEYS						
#B ADENOMA, RENAL TUBULE	0	2	1	1	0.3027	0.4167
#M CARCINOMA, RENAL TUBULE	1	0	0	0	1.0000	1.0000
LIVER						
#B ADENOMA, HEPATOCELLULAR	12	1	7	5	(0.7929)	0.9232
#B ADENOMA, HEPATOCELLULAR, MU	3	0	1	1	(0.7626)	0.8816
#M CARCINOMA, HEPATOCELLULAR	4	3	3	4	(0.3111)	0.4375
LUNGS						
#B ADENOMA, BRONCHIOLO-ALVEOLAR	8	8	12	12	(0.0344)	0.0986
#M CARCINOMA, BRONCHIOLO-ALVEO	4	4	5	4	(0.4049)	0.4604
PANCREAS						
#B ADENOMA, ISLET CELL	1	0	0	0	(1.0000)	1.0000

Table A.4.1. (cont.) Tumorgenicity in Male Mice

PITUITARY						
#M CARCINOMA, PARS DISTALIS	0	0	0	1	(0.2209)	0.4318
SKIN						
#B PAPILLOMA, SQUAMOUS CELL	0	0	0	1	(0.1364)	0.3750
#M SARCOMA, UNDIFFERENTIATED	0	0	1	0	(0.4091)	
SPLEEN						
#M SARCOMA, UNDIFFERENTIATED	0	0	0	1	(0.2391)	0.4400
STOMACH, GLAN						
#M NEUROENDOCRINE CELL TUMOR	0	0	0	1	(0.1395)	0.3750
SYSTEMIC TUMORS						
#B HEMANGIOMA	2	2	5	3	(0.2880)	0.4937
#M HEMANGIOSARCOMA	4	5	5	4	(0.4063)	0.4193
#M LYMPHOMA, MALIGNANT	5	2	2	4	(0.4742)	0.6042
#M MESOTHELIOMA, MALIGNANT	1	0	0	0	(1.0000)	1.0000
#M SARCOMA, HISTIOCYTIC	1	0	2	2	(0.1598)	0.4111
TESTES						
#B ADENOMA, INTERSTITIAL CELL	0	1	0	0	(0.7674)	
THYROID GLANDS						
#B ADENOMA, FOLLICULAR CELL	0	0	0	1	(0.3673)	0.7500
#M CARCINOMA, FOLLICULAR CELL	1	1	0	1	(0.7219)	0.8511
URINARY BLADDER						
#B SUBMUCOSAL MESENCHYMAL TUMOR	0	1	0	0	(0.7674)	

Table A.4.2. Tumorgenicity in Female Mice

Organ / Tumor	Control	Low	Medium	High	p-values: Trend	Hi vs Cntrl
ADRENAL CORTEX						
#B ADENOMA	1	0	0	0	(1.0000)	1.0000
#B ADENOMA, A CELL	1	0	0	1	(0.3442)	0.5183
ADRENAL MEDULLA						
#B PHEOCHROMOCYTOMA, BENIGN	0	1	0	1	(0.3344)	0.7000
BONE						
#M OSTEOSARCOMA	1	0	0	0	(1.0000)	1.0000
BRAIN						
#M SARCOMA, MENINGEAL	0	0	1	0	(0.4813)	
CERVIX						
#B FIBROMA	1	0	1	0	(0.7086)	1.0000
#B GRANULAR CELL TUMOR, BENIGN	0	1	0	0	(0.7500)	
#B LEIOMYOMA	0	1	1	0	(0.7328)	
#M CARCINOMA, SQUAMOUS CELL	1	0	0	0	(1.0000)	1.0000
#M LEIOMYOSARCOMA	1	0	0	0	(1.0000)	1.0000
#M SARCOMA, ENDOMETRIAL STROMAL	0	0	1	0	(0.4722)	
HARDERIAN GLANDS						
#B ADENOMA	1	1	8	3	0.1298	0.4311
#M CARCINOMA	1	1	0	0	0.8810	1.0000
KIDNEYS						
#M RENAL MESENCHYMAL TUMOR,MAL	0	0	1	0	0.5000	

Table A.4.2. (cont.) Tumorgenicity in Female Mice

LIVER						
#B ADENOMA, HEPATOCELLULAR	2	1	0	1	(0.6650)	0.7718
#B LIPOMA	0	0	1	0	(0.4070)	
LUNGS						
#B ADENOMA, BRONCHIOLO-ALVEOLA	6	3	10	5	(0.4381)	0.7670
#M CARCINOMA, BRONCHIOLO-ALVEO	6	1	4	6	(0.1919)	0.4668
MAMMARY GLAND						
#M ADENOCARCINOMA	1	0	1	2	(0.1262)	0.4079
OVARIES						
#B ADENOMA, TUBULOSTROMAL	2	0	0	0	(1.0000)	1.0000
#B CYSTADENOMA	0	1	2	2	(0.0928)	0.2273
#B GRANULOSA CELL TUMOR, BENIGN	1	0	0	0	(1.0000)	1.0000
#B LUTEOMA	1	0	0	0	(1.0000)	1.0000
#B TERATOMA, BENIGN	0	1	0	0	(0.6790)	
#M GRANULOSA CELL TUMOR, MALIG	0	1	0	0	(0.8448)	
#M LEIOMYOSARCOMA	0	0	1	0	(0.4872)	
#M THECOMA, MALIGNANT	0	0	1	0	(0.4321)	
PANCREAS						
#M CARCINOMA, ISLET CELL	0	0	1	0	(0.5000)	
PITUITARY						
#B ADENOMA, PARS DISTALIS	1	2	1	0	(0.8302)	1.0000
SKIN						
#B PILOMATRICOMA	0	0	0	2	(0.0345)	0.1371
#M SARCOMA, UNDIFFERENTIATED	0	1	1	1	(0.2630)	0.4444
STERNUM						
#M SARCOMA, UNDIFFERENTIATED	0	1	0	0	(0.7521)	
STOMACH, GLAN						
#B POLYP	1	0	0	0	(1.0000)	1.0000
STOMACH, NON						
#B PAPILLOMA, SQUAMOUS CELL	0	0	1	0	(0.4070)	
SYSTEMIC TUMORS*						
#B HEMANGIOMA	9	5	3	6	(0.9356)	0.9307
#M FIBROUS HISTIOCYTOMA, MALIG	0	0	0	1	(0.1803)	0.3667
#M HEMANGIOSARCOMA	4	5	6	6	(0.2499)	0.2536
#M LEUKEMIA, GRANULOCYTIC	0	2	0	0	(0.7671)	
#M LYMPHOMA, MALIGNANT	13	10	8	12	(0.2019)	0.1862
#M SARCOMA, HISTIOCYTIC	5	9	9	4	(0.5363)	0.7387
URINARY BLADDER						
#B SUBMUCOSAL MESENCHYMAL TUMOR	0	1	0	0	(0.7750)	
UTERUS						
#B GRANULAR CELL TUMOR, BENIGN	0	1	0	0	(0.7750)	
#B LEIOMYOMA	1	1	1	1	(0.4959)	0.7749
#B POLYP, ENDOMETRIAL STROMAL	4	7	5	6	(0.3682)	0.3452
#B POLYP, ENDOMETRIAL STROMAL,	0	0	1	0	(0.4070)	
#M CARCINOMA	3	2	2	1	(0.8639)	0.9338
#M LEIOMYOSARCOMA	3	2	0	0	(0.9945)	1.0000
#M SARCOMA, ENDOMETRIAL STROMAL	1	1	1	0	(0.7160)	1.0000

* The Sponsor indictes that all animals were examined for these systemic tumors, but since not all organs were examined, this claim does not seem to justifisble.

Table A.4.3. Tumorgenicity in Male Rats

Organ / Tumor	Control	Low	Medium	High	p-values: Trend	Hi vs Cntrl
ADRENAL MEDULLA						
#B PHEOCHROMOCYTOMA, BENIGN	2	5	5	5	(0.2436)	0.2869
#M PHEOCHROMOCYTOMA, MALIGNANT	0	2	1	3	(0.1193)	0.1515
AORTA						
#M HIBERNOMA, MALIGNANT	0	0	1	0	(0.5444)	
BRAIN						
#M ASTROCYTOMA, MALIGNANT	0	1	0	2	(0.1055)	0.2406
BULBOURETHRAL GL						
#M FIBROSARCOMA	0	0	0	1	(0.2000)	0.3913
FEMUR						
#M OSTEOSARCOMA	0	0	0	1	(0.2688)	0.5375
KIDNEYS						
#M LIPOSARCOMA	0	1	0	0	0.7727	
LIVER						
#B ADENOMA, HEPATOCELLULAR	0	2	0	1	(0.4923)	0.5435
#M CARCINOMA, HEPATOCELLULAR	0	1	1	0	(0.6694)	
LUNGS						
#M HIBERNOMA, MALIGNANT	0	2	0	0	(0.8335)	
LYMPH NODE, AXI						
#B LIPOMA	0	0	0	1	(0.2778)	0.5435
MAMMARY GLAND						
#B ADENOMA	0	0	1	0	(0.6429)	
#B FIBROADENOMA	0	0	2	0	(0.5049)	
#M ADENOCARCINOMA	1	0	1	0	(0.8241)	1.0000
ORAL CAVITY						
#M OSTEOSARCOMA	0	0	0	1	(0.2778)	0.5435
PANCREAS						
#B ADENOMA, ISLET CELL	4	0	2	7	(0.0432)	0.2722
#M CARCINOMA, ISLET CELL	1	0	0	1	(0.5619)	0.8081
PARATHYROIDS						
#B ADENOMA	1	0	0	0	(1.0000)	1.0000
PITUITARY						
#B ADENOMA, PARS DISTALIS	27	28	27	22	(0.8950)	0.8625
RECTUM						
#M LEIOMYOSARCOMA	1	0	0	0	(1.0000)	1.0000
SEMINAL VESICLES						
#M CARCINOMA, UNDIFFERENTIATED	0	1	0	0	(0.7727)	
SKIN						
#B BASAL CELL TUMOR, BENIGN	1	0	1	2	(0.2758)	0.6020
#B FIBROMA	1	1	5	1	(0.5020)	0.7971
#B KERATOACANTHOMA, BENIGN	0	2	1	0	(0.7843)	
#B LIPOMA	2	0	2	0	(0.8215)	1.0000
#B PAPILLOMA, SQUAMOUS CELL	0	0	0	1	(0.2809)	0.5435
SOFT TISSUE- THO						
#B HIBERNOMA	0	0	0	1	(0.2759)	0.6154
#M HIBERNOMA, MALIGNANT	2	1	0	1	(0.7929)	0.8724
STOMACH, NON						
#B PAPILLOMA, SQUAMOUS CELL	1	0	0	0	(1.0000)	1.0000

Table A.4.3. (cot.) Tumorgenicity in Male Rats

SYSTEMIC TUMORS						
#M LEUKEMIA, GRANULOCYTIC	0	0	0	1	(0.2759)	0.6154
#M LYMPHOMA, MALIGNANT	1	0	2	0	(0.7800)	1.0000
#M SARCOMA, HISTIOCYTIC	1	3	2	2	(0.4667)	0.5023
TAIL						
#B KERATOACANTHOMA, BENIGN	0	0	0	1	(0.2759)	0.6154
#B PAPILLOMA, SQUAMOUS CELL	0	0	1	0	(0.5444)	
TESTES						
#B ADENOMA, INTERSTITIAL CELL	2	0	1	2	(0.4935)	0.7794
THYMUS						
#M FIBROSARCOMA	1	0	0	0	(1.0000)	1.0000
THYROID GLANDS						
#B ADENOMA, C-CELL	4	3	6	2	(0.6809)	0.8969
#B ADENOMA, FOLLICULAR CELL	0	0	2	0	(0.3619)	
#M CARCINOMA, C-CELL	1	0	0	2	(0.2561)	0.5849
URINARY BLADDER						
#B LEIOMYOMA	1	0	0	0	(1.0000)	1.0000

Table A.4.4. Tumorgenicity in Female Rats

Organ / Tumor	Control	Low	Medium	High	p-values: Trend	Hi vs Cntrl
ADIPOSE TISSUE						
#M HIBERNOMA, MALIGNANT	0	1	0	0	(0.7273)	
ADRENAL CORTEX						
#B ADENOMA	1	1	1	0	(0.8395)	1.0000
ADRENAL MEDULLA						
#B PHEOCHROMOCYTOMA, BENIGN	0	0	0	1	(0.2368)	0.4500
#M PHEOCHROMOCYTOMA, MALIGNANT	1	1	1	0	(0.8297)	1.0000
BRAIN						
#M ASTROCYTOMA, MALIGNANT	0	0	1	1	(0.1683)	0.4500
#M RETICULOSIS, MALIGNANT	0	0	1	0	(0.5152)	
CERVIX						
#B LEIOMYOMA	0	1	0	0	(0.7105)	
#B PAPILLOMA, SQUAMOUS CELL	0	0	1	0	(0.5278)	
#B POLYP, ENDOMETRIAL STROMAL	0	0	1	0	(0.4605)	
#M SARCOMA, ENDOMETRIAL STROMAL	0	1	0	0	(0.7105)	
CLITORAL GL						
#B PAPILLOMA, SQUAMOUS CELL	0	1	0	1	(0.2706)	0.4359
EAR(S)						
#B PAPILLOMA, SQUAMOUS CELL	1	0	0	1	(0.4829)	0.7088
HEART						
#B SCHWANNOMA, ENDOCARDIAL, BE	1	0	0	0	(1.0000)	1.0000
ILEUM						
#B LEIOMYOMA	0	0	1	0	(0.5000)	
JEJUNUM						
#B LEIOMYOMA	1	0	0	0	(1.0000)	1.0000

Table A.4.4. (cont.) Tumorgenicity in Female Rats

KIDNEYS						
#B ADENOMA, RENAL TUBULE	0	0	1	0	0.4483	
#B LIPOMA	0	0	0	1	0.2368	0.4500
LUNGS						
#M CARCINOMA, BRONCHIOLO-ALVEO	0	1	0	0	(0.7222)	
LYMPH NODE, MED						
#M LIPOSARCOMA	0	0	1	0	(0.4861)	
MAMMARY GLAND						
#B ADENOMA	3	3	3	2	(0.6693)	0.7694
#B FIBROADENOMA	18	15	9	17	(0.5625)	0.5830
#M ADENOCARCINOMA	12	14	7	12	(0.5431)	0.4611
#M CARCINOMA, UNDIFFERENTIATED	0	0	0	2	(0.0517)	0.2022
#M FIBROSARCOMA	2	0	0	0	(1.0000)	1.0000
NASAL LEVEL III						
#M CARCINOMA, SQUAMOUS CELL	1	0	0	0	(1.0000)	1.0000
OVARIES						
#B LUTEOMA	1	0	0	0	(1.0000)	1.0000
#B SERTOLI CELL TUMOR, BENIGN	0	0	1	0	(0.7143)	
#M MESOTHELIOMA, MALIGNANT	0	1	0	0	(0.7105)	
PANCREAS						
#B ADENOMA, ISLET CELL	2	0	0	0	(1.0000)	1.0000
#M CARCINOMA, ACINAR CELL	0	1	0	0	(0.7188)	
PITUITARY						
#B ADENOMA, PARS DISTALIS	40	34	36	36	(0.6689)	0.6793
SKIN						
#B BASAL CELL TUMOR, BENIGN	0	1	0	0	(0.7067)	
#B LIPOMA	0	1	0	1	(0.2822)	0.4500
#M BASAL CELL TUMOR, MALIGNANT	0	0	0	1	(0.2400)	0.4500
#M SCHWANNOMA, MALIGNANT	0	0	0	1	(0.2400)	0.4500
SOFT TISSUE- THO						
#M HIBERNOMA, MALIGNANT	1	0	0	0	(1.0000)	1.0000
SYSTEMIC TUMORS						
#M HEMANGIOSARCOMA	0	0	1	0	(0.5152)	
#M LEUKEMIA, GRANULOCYTIC	2	1	2	3	(0.2789)	0.4678
#M LYMPHOMA, MALIGNANT	0	0	0	1	(0.2252)	0.4359
#M SARCOMA, HISTIOCYTIC	1	0	0	0	(1.0000)	1.0000
THYMUS						
#B THYMOMA, BENIGN	0	0	1	0	(0.4605)	
#M HIBERNOMA, MALIGNANT	0	0	1	1	(0.1703)	0.4636
#M LIPOSARCOMA	0	1	0	0	(0.8125)	
THYROID GLANDS						
#B ADENOMA, C-CELL	2	2	1	4	(0.1871)	0.2321
#B ADENOMA, FOLLICULAR CELL	2	1	0	0	(0.9856)	1.0000
#M CARCINOMA, C-CELL	0	0	2	0	(0.4756)	
#M CARCINOMA, FOLLICULAR CELL	0	1	0	0	(0.7105)	
UTERUS						
#B ADENOMA, ENDOMETRIAL	0	0	1	0	(0.4605)	
#B POLYP, ENDOMETRIAL STROMAL	3	4	0	4	(0.3949)	0.4058
#M CARCINOMA	0	0	2	1	(0.1666)	0.4321
#M SCHWANNOMA, MALIGNANT	1	0	0	0	(1.0000)	1.0000
VAGINA						
#B LEIOMYOMA	0	0	0	1	(0.2571)	0.4737

Appendix 5. 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.

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
5/31/2006 03:13:31 PM
BIOMETRICS

Karl Lin
6/2/2006 08:35:55 AM
BIOMETRICS
Concur with review