CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

202107Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA Serial Number: 202107

Drug Name: Corlux® (mifepristone)

Indication(s): To reduce the effects of hypercortisolism in patients with endogenous Cushing's syndrome

Applicant: Corcept Therapeutics

Date(s): 04/25/2011

Review Priority: Standard

Biometrics Division: 2

Statistical Reviewer: Japobrata Choudhury, Ph.D.

Concurring Reviewers: Todd Sahlroot, Ph.D.

Medical Division: Metabolic and Endocrine Products

Clinical Team: Marina Zemskova, M.D., Dragos Roman, M.D., Mary Parks, M.D.

Project Manager: Jena Weber

Keywords:

Link to DBII key Words: Clinical Studies, NDA Review

http://intranetapps.fda.gov/scripts/ob_apps/ob/eWork/uploads/eWork/2009/Keywords-in-DFS.htm

Table of Contents

1. EXECUTIVE SUMMARY

The objective of this study was to examine the safety and efficacy of mifepristone for treatment of the signs and symptoms of hypercortisolemia in subjects with endogenous Cushing's syndrome from ACTH-dependent or ACTH-independent disorders.

This was a 24-week, open-label study of the administration of mifepristone to subjects with Cushing's syndrome. The sponsor states, "An open-label design was chosen for this study because of the lack of an approved comparator drug that was available commercially." Following a screening period of up to 6 weeks, 50 subjects were assigned to receive 300 mg mifepristone once daily (QD). Because the optimal dose of mifepristone for each subject was not known, dose escalation was undertaken cautiously with careful observation of clinical status. Dose escalations beyond 300 mg were made under some conditions.

Subjects belonged to one of two study cohorts. The C-DM cohort (n=29) consisted of subjects with Cushing's syndrome and diabetes or impaired glucose tolerance. The C-HT cohort (n=21) consisted of subjects with Cushing's syndrome and a diagnosis of hypertension only (without diabetes or impaired glucose tolerance). Each cohort had a separate primary efficacy endpoint.

The primary endpoint for subjects in **C-DM** was evaluation of response based on the change in AUC for glucose ($AUC_{glucose}$) from baseline to Week 24/ET for the 2-hour oGTT in the mITT population. A responder was a subject who had at least a 25% decrease in AUC from baseline. A response in $AUC_{glucose}$ was observed in 60% of the subjects (1-sided 95% CI lower bound, 42%). The sponsor considered this result to be statistically significant because the lower bound of the 95% CI was greater than 20%, the pre-specified margin of clinical significance. I also computed a 2-sided 95% confidence interval for the response rate. The lower bound of the 2-sided 95% confidence interval was 40.4%. The mean change from baseline in AUC was -8722 (2-sided 95% CI = (-13184, -4260), p=.0009) from a baseline mean of 30670.

	C-DM N=25 n (%)	**Lower Bound 1-sided 95% Exact Binomial CI**
Subjects who had at least a 25% decrease from baseline in $AUC_{glucose}$ at Week 24/ET		
Responder	15 (60.0)	41.68%
Non-responder	10 (40.0)	

HbA1c was not the primary endpoint in C-DM (it was a secondary endpoint) but nevertheless is an important clinical measure of diabetic control. The mean change from baseline in HbA1c was -1.11 (2-sided 95% CI = (-1.56, -0.65), p=.0001) from a baseline mean of 7.36. While it can be difficult to assess changes from baseline in $AUC_{glucose}$ and HbA1c in the absence of a control group, the observed changes were of sufficient magnitude so that they could be attributed to the action of the drug since hyperglycemia would be expected to persist without treatment and in the absence of significant fluctuations in cortisol and ACTH levels.

Nevertheless, clinical judgment should be given priority in this open-label study with titration and meager data.

The primary efficacy variable for subjects in **C-HT** was evaluation of response based on the change in diastolic blood pressure from baseline to Week 24. A responder was a subject who had at least a 5 mmHg reduction in dBP from baseline. A response for diastolic blood pressure was observed in 38% of the subjects (1-sided 95% CI lower bound, 21%). The sponsor considered this result to be statistically significant because the lower bound of the 95% CI was greater than 20%, the pre-specified margin of clinical significance. I also computed a 2-sided 95% confidence interval for the response rate. The lower bound of the 2-sided 95% confidence interval was 16.8% which fell below the margin.

	C-HT N=21 n (%)	**Lower Bound 1-sided 95% Exact Binomial CI**
Subjects who had at least a 5 mmHg reduction from baseline in diastolic blood pressure at Week 24/ET		
Responder	8 (38.1)	20.57%
Non-responder	13 (61.9)	

The mean change from baseline in dBP (mmHg) was -0.1 (2-sided 95% CI = (-4.6, 4.6), p=.98) from a baseline mean of 82.9. Therefore, across the two dBP endpoints, there was no statistical evidence of diastolic blood pressure lowering in the C-HT cohort.

Labeling

Though no statistical significance was claimed for the secondary efficacy variables, there is one danger that non-statisticians may not be fully alert that these descriptive statistics do not mean much. They are just numerical results based on one sample; there is no assurance or confidence regarding the population or the reality.

The definition of a Responder in the key secondary efficacy variable: "A responder was defined as a subject whose median reviewer score was + 1 at any reviewed visit after baseline through Week 24/ET" with the phrase "at any reviewed visit," gives multiple opportunities for a success and is not as dependable as a response at any one time-point. Therefore, the results of this key variable, if are allowed to be in the labeling at all, this point should be emphasized.

2. INTRODUCTION

2.1 Overview

Note: Tables and Figures presented in this document are referenced by "below" or "above". Those referenced with an extended numbering system are in the NDA Study Report. Unless mentioned otherwise, the source of all information is the sponsor's submission. The reviewer's interpretations, comments, or conclusions are clearly identified under notes, comments, or separate sections.

Study Title:	An Open-label Study of the Efficacy and Safety of CORLUX® (mifepristone) in the Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome
Study Number:	C1073-400
Study Phase:	III
Study Design:	open label, uncontrolled
Product Name:	Mifepristone
Formulation:	tablet
Indication:	Endogenous Cushing's Syndrome
Study Initiated (first subject enrolled):	12 August 2008
Study Completed (last subject completed):	10 January 2011
Principal Investigators:	Multicenter (see Appendix 16.1.4)
Final Date:	31 March 2011

Objectives: The primary objective of the study was to evaluate the safety and efficacy of mifepristone in the treatment of the signs and symptoms of endogenous Cushing's syndrome.

Methodology: This was a 24-week, open-label study of the administration of mifepristone to subjects with Cushing's syndrome. Following a screening period of up to 6 weeks, 50 subjects were assigned to receive 300 mg mifepristone once daily (QD). Because the optimal dose of mifepristone for each subject was not known, dose escalation was undertaken cautiously with careful observation of key signs and symptoms of Cushing's syndrome. Dose escalations beyond 300 mg were made under the following conditions:

If no clinical improvement had been seen,

If the drug had been well tolerated, and

Based on the subject's weight at the escalation visit.

After 14 days of dosing at 300 mg QD, the dose of mifepristone could have been increased as outlined below:

Subject Weight	Day 1	Day 14	Week 6	Week 10[a]
< 60 kg	300 mg	600 mg	900 mg	900 mg
≥ 60 kg	300 mg	600 mg	900 mg	1200 mg

All doses were given as a single daily dose.
a Dose escalation stopped at Week 6 for subjects weighing <60 kg.

Dose escalation was not required if significant clinical improvement was noted at the current dosing level. In cases of severe hypercortisolism the dose of mifepristone could have been increased beyond 1200 mg QD (or 900 mg QD for subjects weighing -: 60 kg) with the approval of the medical monitor (note: there was no such occurrence in this study). However, the dose was not to be increased beyond a weight-adjusted dose of 20 mg/kg per day. Subjects who completed this study (CI073-400) were given the opportunity to continue receiving mifepristone by entering an extension study under a separate protocol (Study C1073-415). Number of Subjects (Planned and Analyzed): 50 subjects were planed; 50 subjects were analyzed for safety; 46 subjects were analyzed for efficacy in the modified intent-to-treat (mITT) population. There were 29 subjects in the cohort of subjects with diabetes mellitus and/or impaired glucose tolerance (C-DM) and 21 subjects in the cohort of subjects with hypertension (C-HT).

Diagnosis and Main Criteria for Inclusion: Male or female subjects who were >=18 years old with endogenous Cushing's syndrome and had Type 2 diabetes (or impaired glucose function) OR hypertension were eligible to participate in this study.

Test Product, Dose and Mode of Administration, Lot Number: 300 mg mifepristone tablets administered orally once daily; number of tablets adjusted depending upon escalation schedule described above.

Duration of Treatment: up to 24 weeks

Criteria for Evaluation: Efficacy:
The primary efficacy endpoints (described in further detail below) were based on assessments of glucose and blood pressure.

The key secondary efficacy endpoint was based on Data Review Board (DRB) assessments of the signs and symptoms Cushing's syndrome as well as laboratory findings.

Statistical Methods:

Efficacy: Primary Efficacy Endpoints:

The primary endpoint for subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance) (C-DM cohort) was the change in the area under the concentration-time cure for glucose (AUCglucose) in the 2-hour oral glucose tolerance test (oGTT) from baseline to

Week 24 in subjects with diabetes/impaired glucose tolerance with or without hypertension at screening. A responder analysis using the modified Intent-to- Treat (mITT) population was used to measure success on this primary efficacy endpoint. A responder was defined as a subject who experienced at least a 25% decrease in AUCglucose from baseline to Week 24. This efficacy measurement was to be declared successful if the lower limit of the exact 95% binomial confidence interval for the responder rate was >= 20%.

Changes in diastolic blood pressure were analyzed for subjects with Cushing's syndrome and a diagnosis of hypertension (C-HT) at screening (i.e., without impaired glucose tolerance or diabetes) as a primary endpoint. A responder analysis of the reduction in diastolic blood pressure from baseline to Week 24 was performed for the efficacy population. A responder was defined as a subject who experienced a >= 5 mmHg decline in diastolic blood pressure from baseline to Week 24. This efficacy measurement was to be declared successful if the lower limit of the exact 95% binomial confidence interval for the responder rate was >=20%.

The study was considered to have had a positive outcome and to have achieved the primary endpoint if either of the two primary measures described above were positive.

Key Secondary Endpoint:

Clinical improvement as determined by the DRB: a responder analysis was used to determine clinical improvement in all subjects. The DRB performed a review of eight categories of clinical parameters to evaluate whether a subject's signs and symptoms of Cushing's syndrome had changed. For this secondary efficacy endpoint, a responder was defined as a subject whose median reviewer score was +1 (of possible ratings of -1, 0, or + 1) at any reviewed visit after baseline though Week 24/ET. This efficacy measurement was to be declared successful if the lower limit of the exact 95% binomial confidence interval for the responder rate was >= 30%.

----------**INDICATIONS AND USAGE**---------- (b) (4) (mifepristone) is a cortisol receptor blocker indicated to treat the clinical and metabolic effects of hypercortisolism in patients with endogenous Cushing's syndrome, including:

- Patients with Cushing's disease who have not adequately responded to or relapsed after surgery
- Patients with Cushing's disease who are not candidates for surgery

(b) (4)

Mifepristone (C-I073, referred to previously as CORLUX) is a glucocorticoid-receptor type II antagonist (GR-II) that also acts as an antagonist of the progesterone receptor.

Mifepristone, in a 200 mg tablet formulation, is currently approved and marketed in the United States (US) as Mifeprex® by Danco Laboratories, LLC for the medical termination of pregnancy. (b) (4)

2.2 Data Sources

\\fdswa150\NONECTD\N202107\S_001\2011-04-15\Folder 6\n202107\m5\datasets\c1073-400

3. STATISTICAL EVALUATION

3.1 Data and Analysis Quality

No apparent concern.

3.2 Evaluation of Efficacy

Study Design and Endpoints

This was a 24-week, open-label study of the administration of mifepristone to subjects with Cushing's syndrome. Following a screening period of up to 6 weeks, 50 subjects were assigned to receive 300 mg mifepristone once daily (QD). Because the optimal dose of mifepristone for each subject was not known, dose escalation was undertaken cautiously with careful observation of clinical status. Dose escalations beyond 300 mg were made under the following conditions:

If no clinical improvement had been seen, If the drug had been well tolerated, and Based on the subject's weight at the escalation visit.

After 14 days of dosing at 300 mg QD, the dose of mifepristone could have been increased as outlined in Table below:

Subject Weight	Day 1	Day 14	Week 6	Week 10[a]
< 60 kg	300 mg	600 mg	900 mg	900 mg
≥ 60 kg	300 mg	600 mg	900 mg	1200 mg

All doses were given as a single daily dose.
a Dose escalation stopped at Week 6 for subjects weighing <60 kg.

Dose escalation was not required if significant clinical improvement was noted at the current dosing level. In cases of severe hypercortisolism, the dose of mifepristone could have been increased beyond 1200 mg QD (or 900 mg QD for subjects weighing 0: 60 kg) with the approval of the medical monitor; however, the dose was not to be increased beyond a weight-adjusted dose of 20 mg/kg per day. Dosing was to be interrupted and the subjects treated with

exogenous corticosteroids if signs of adrenal insufficiency were noted.

The primary efficacy assessments were based on measurements of glucose (based on 2-hour, 75 gram oral glucose tolerance tests (oGTT) and blood pressure.

Other secondary and exploratory efficacy assessments included:

use of concomitant medications for diabetes and hypertension
body weight
hemoglobin HbA1c (glycosylated hemoglobin)
systolic blood pressure
skin and physical appearance (including photographs)
waist circumference
body composition
bone density of the spine and hip (based on dual-energy x-ray absorptiometr (DXA) scan)
bone metabolism markers (osteocalcin, urinary N-telopeptide of type I collagen (NTxJ, and bone specific alkaline phosphatase)
cognitive and psychiatric assessments (Beck Depression Inventory n and Trail Making Test); and the Short Form (SF)-36 Health Survey.
muscle strength (sit-to-stand test and hand grip test)
insulin levels
laboratory measurements of thrombin-antithrombin (TAT), e-selectin, and adiponectin.

Patient Disposition, Demographic and Baseline Characteristics

Disposition of Subjects

The disposition of the study subjects is presented in Table below. A total of 50 subjects were enrolled in the study, and 34 completed the study. Of the 16 subjects who withdrew from the study, seven withdrew because of AEs (including one who died subsequently due to underlying illness) and two subjects died due to underlying illness while still participating in the study; these subjects are described in further detail in Section 12.3.1.3 of the NDA Report. Five subjects withdrew consent, one subject was too ill to travel (and subsequently died due to underlying illness) and one subject was withdrawn due to non-compliance with study procedures.

Overall, 34 (68%) of the 50 subjects completed the 24-week treatment period (20 of 29 subjects in the C-DM cohort and 14 of 21 subjects in the HT cohort). In total, 40 subjects (80%) attended the 6-week follow-up visit (22 in the C-DM cohort and 18 in the C-HT cohort), including the 34 subjects who completed the 24-week treatment period as well as six subjects who terminated early.

Subject Disposition (ITT/Safety Population)

Number of	C-DM (N=29) n (%)	C-HT (N=21) n (%)	Overall (N=50) n (%)
Subjects screened			84
Screen failures			34
Subjects who completed the study	20 (69.0)	14 (66.7)	34 (68.0)
Subjects who withdrew	9 (31.0)	7 (33.3)	16 (32.0)
Reasons for withdrawal			
Adverse event	2 (6.9)	5 (23.8)	7 (14.0)
Death	1 (3.4)	1 (4.8)	2 (4.0)
Withdrawal by subject	4 (13.8)	1 (4.8)	5 (10.0)
Other[a]	2 (6.9)	0 (0.0)	2 (4.0)

C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); C-HT = subjects with Cushing's syndrome and hypertension; oGTT = oral glucose tolerance test.

The C-DM group includes subjects with diabetes mellitus type 2 and/or impaired glucose tolerance at screening and Day 1 as determined by 2 or more abnormal oGTTs. The C-HT group includes subjects with a diagnosis of hypertension at screening but without diabetes mellitus type 2 and/or impaired glucose tolerance. Three subjects in the hypertension group entered with a diagnosis of diabetes not confirmed by 2 abnormal oGTTs.

The ITT/Safety population is defined as all enrolled subjects who received at least 1 dose of study medication.

a Subject 07-008 withdrew because she was too ill to travel for required study visits and Subject 08-013 is listed as "withdrawal by subject due to severe non-compliance" (see Listing 16.2.1.1).

Source: Table 14.1.1.1.

Demographics and Other Baseline Characteristics

Demographics and body measurements of the study population at baseline are summarized in Table below. The majority of subjects were female (35/50, 70%) and white (42/50, 84%); the mean age of the study population was 45.4 years. Overall, the mean of weight, BMI, and waist circumference measurements at baseline were 99.5 kg, 35.7 kg/m2, and 119 cm, respectively.

Demographics and Body Measurements at Baseline (ITT/Safety Population)

Characteristic	C-DM (N=29)	C-HT (N=21)	Overall (N=50)
Sex, n (%)			
Male	7 (24.1)	8 (38.1)	15 (30.0)
Female	22 (75.9)	13 (61.9)	35 (70.0)
Race, n (%)			
Black or African American	6 (20.7)	2 (9.5)	8 (16.0)
White	23 (79.3)	19 (90.5)	42 (84.0)
Ethnicity, n (%)			
Hispanic or Latino	2 (6.9)	2 (9.5)	4 (8.0)
Not Hispanic or Latino	27 (93.1)	19 (90.5)	46 (92.0)
Age, years			
Mean (SD)	44.4 (13.71)	46.7 (8.83)	45.4 (11.85)
Median	41.0	46.0	45.0
Min, Max	26, 71	26, 67	26, 71
Height, cm			
Mean (SD)	168 (12.11)	166 (8.84)	167 (10.81)
Median	168	163	166
Min, Max	143.5, 190.5	154.0, 185.4	143.5, 190.5
Weight, kg			
Mean (SD)	105 (33.54)	91.4 (21.10)	99.5 (29.55)
Median	102	88.2	92.4
Min, Max	61.3, 198.7	62.7, 150.5	61.3, 198.7
BMI, kg/m^2			
Mean (SD)	37.4 (11.18)	33.4 (7.44)	35.7 (9.90)
Median	35.1	31.8	33.5
Min, Max	24.1, 66.4	24.5, 53.6	24.1, 66.4
Waist circumference, cm			
Mean (SD)	124 (21.73)	111 (15.77)	119 (20.31)
Median	120	104	115
Min, Max	97.9, 178.4	88.5, 153.5	88.5, 178.4

BMI = body mass index; C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); C-HT = subjects with Cushing's syndrome and hypertension; ITT = intent-to-treat; Max = maximum; Min = minimum; oGTT = oral glucose tolerance test; SD = standard deviation.

The C-DM group included subjects with diabetes mellitus type 2 and/or impaired glucose tolerance at screening and Day 1 as determined by 2 or more abnormal oGTTs. The C-HT group included subjects with a diagnosis of hypertension at screening but without diabetes mellitus type 2 and/or impaired glucose tolerance. Three subjects in the hypertension group (Subjects 07-006, 07-007, and 24-002) entered with a diagnosis of diabetes not confirmed by two abnormal oGTTs.

The ITT/Safety population is defined as all enrolled subjects who received at least 1 dose of study medication.

Source: Table 14.1.2.1.

Cushing's Syndrome History

Table below presents Cushing's syndrome etiology and history as well as the signs and symptoms of Cushing's syndrome noted at the screening visit.

A total of 43 subjects had Cushing's disease as the etiology for their Cushing's syndrome. Four subjects had Cushing's syndrome with an etiology of ectopic ACTH secretion, and three subjects had Cushing's syndrome with an etiology of adrenal carcinoma. All subjects with Cushing's disease except one (Subject 24-005) had previously undergone pituitary surgery (Listing 16.2.4.2.2).

Cushing's Syndrome History and Signs/Symptoms at Screening (ITT/Safety Population)

Characteristic	C-DM (N=29) n (%)	C-HT (N=21) n (%)	Overall (N=50) n (%)
Cushing's etiology/syndrome history			
Cushing's disease	24 (82.8)	19 (90.5)	43 (86.0)
Ectopic (ACTH)	3 (10.3)	1 (4.8)	4 (8.0)
Adrenal carcinoma	2 (6.9)	1 (4.8)	3 (6.0)
Ectopic CRH secretion	0 (0.0)	0 (0.0)	0 (0.0)
Adrenal adenoma	0 (0.0)	0 (0.0)	0 (0.0)
Adrenal autonomy	0 (0.0)	0 (0.0)	0 (0.0)
Cushing's disease that			
Recurred after primary pituitary surgery	12 (41.4)	9 (42.9)	21 (42.0)
Persisted despite pituitary surgery (failed pituitary surgery)	16 (55.2)	14 (66.7)	30 (60.0)
Had been treated with radiation therapy to the pituitary	12 (41.4)	6 (28.6)	18 (36.0)
Was not treatable with surgery[a]	5 (17.2)	1 (4.8)	6 (12.0)
Existed in subjects who were not candidates for surgery[a]	5 (17.2)	1 (4.8)	6 (12.0)
Subjects with diabetes mellitus or impaired glucose tolerance at screening, n (%)	29 (100.0)	0 (0.0)	29 (58.0)
Subjects with hypertension at screening, n (%)[b]	22 (75.9)	21 (100.0)	43 (86.0)
Signs and symptoms related to hypercortisolemia, n (%)			
Cushingoid appearance	29 (100.0)	20 (95.2)	49 (98.0)
Hirsutism and/or violaceous striae or acne	19 (65.5)	11 (52.4)	30 (60.0)
Increased body weight or central obesity	29 (100.0)	20 (95.2)	49 (98.0)
Proximal muscle weakness	15 (51.7)	12 (57.1)	27 (54.0)
Low bone mass (DXA T-score < -1.0)	3 (10.3)	10 (47.6)	13 (26.0)
Psychiatric symptoms (depression or psychosis)	17 (58.6)	9 (42.9)	26 (52.0)

ACTH = adrenal corticotropic hormone; C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); C-HT = subjects with Cushing's syndrome and hypertension; CRH = corticotropin-releasing hormone; DXA = dual-energy x-ray absorptiometry; ITT = intent-to-treat; .

Items under Cushing's disease are not mutually exclusive.

a Only one subject had *de novo* Cushing's disease but was not a candidate for surgery (24-005). Other subjects who had Cushing's disease that "was not treatable with surgery" or "existed in subjects who were not candidates for surgery" had previously undergone pituitary surgery.

b "Subjects with hypertension at screening" category is based on a checkbox on the Inclusion Criteria case report form page.

Source: Table 14.1.2.1 and Table 14.1.4.1.

Statistical Methodologies

The following is copied from the Statistical Analysis Plan, dated Nov 5, 2010:

"Primary Endpoint

Two separate primary efficacy endpoints will be assessed in study subjects with Cushing's syndrome based upon their co-morbid diagnosis of either Diabetes Mellitus Type 2 and/or impaired glucose tolerance or co-morbid hypertension (without co-existing Diabetes Mellitus Type 2 and/or impaired glucose tolerance as follows:

C-DM: Change in glucose tolerance as measured by $AUC_{glucose}$ on 5 point, 2 hour, 75 gram oral glucose tolerance tests.

C-HT: Reduction in diastolic blood pressure.

Key Secondary Endpoint

For both C-DM and C-HT subjects, the key secondary endpoint will be an overall assessment of change in clinical status in Cushing's sign and symptoms, and laboratory findings. These efficacy assessments at each of the key tie points will be conducted for each subject by the Data Review Board (DRB), an independent panel of expert reviewers with expertise in Cushing's Syndrome. The reviewers will be blinded to the dates and sequence of all visits except the baseline and the 6-week safety follow-up visit. (Refer to Section 9.2.1 for description of analysis.)

…

5.5.2 Treatment Duration

Total duration of treatment is calculated as the difference between the dates of last and first dose of study medication plus one day. These dosing dates will be obtained from the Drug Diary data. The first trial dose date (day 1) and the last trial dose date are defied as the dates on which the first dose and last doses are taken, respectively, as shown in the Drug Diary log. All subjects who have completed a total of 30 days of dosing, whether or not those 30 days were contiguous, will be regarded as part of the efficacy population. As par of the validation procedure, programmers will create a validation program that will independently verify that all days on drug for each subject have been included in the tabulation.

5.6 Linear Trapezoidal Rule

The following parameters will be calculated based on the glucose and insulin concentration values from the 2-hour (5-point) oral glucose tolerance test (oGTT):

$AUC_{glucose(0-120)}$ = area under the glucose concentration curve from time 0 to 120 minutes

$AUC_{insulin(0-120)}$ = area under the insulin concentration curve from time 0 to 120 minutes

All AUC calculations will be performed using SAS version 9.2. These calculations will be validated using hand calculation. (see section 11.1.1 for expanded formula for hand calculations). Concentrations will not be corrected for baseline (time 0).

Missing data for $AUC_{glucose}$ will be handled as described in Section 7.4.

6.1 Modified-Intent-to- Treat Population

The efficacy population is defined as all subjects who received a total of at least 30 days of mifepristone during the 24-week treatment period. The primary and secondary endpoints will be evaluated using this modified Intent-to-treat population (mITT). The 30 days of treatment do not need to be consecutive. Subjects who receive >=30 days of mifepristone but terminate prior to week 24 are included in the mITT population. The primary endpoint and secondary efficacy analyses will exclude the following subjects:

C-DM:
Did not undergo at least one oGTT at baseline (day 1) or did not have any oGTT with a valid $AUC_{glucose}$ measurement after day 1 (see Section 5.6 for the definition of valid $AUC_{glucose}$ measurement).

C-HT:
Did not have blood pressure measured (or measured incorrectly) at baseline (day 1) or did not have any valid measurement after day 1.

ALL SUBJECTS:
Entered the study with non-endogenous Cushing's syndrome.

6.2 Completer Population
The study completer population will consist of all subjects (C-DM and C-HT) who complete though the Week 24 visit, are on study drug at the tie of the week 24 visit, and have been compliant with study medication. Compliance for each subject will be defined as having taken at least 80% of the study medication doses as described in the protocol.

…

7.2 Site

There will be up to 30 clinical centers recruiting subjects. It is expected that many sites may enter less than two subjects. Site will not be included in the analysis of the primary endpoint due to the small sample size.

Potential site effects will be explored by summarizing primary analyses at relatively small versus relatively large sites, using the median number of subjects per site to define small and large.

7.3 Sample Size

The sample size of 50 subjects was chosen by clinical judgment. Published case reports suggest that mifepristone can reverse many of the signs and symptoms of hypercortisolemia. However, there have been no prospective clinical trials from which to estimate a treatment effect or standard deviation of treatment effect in subjects with Cushing's Syndrome.

Prior to Amendment 5, the protocol stated that sample size would be re-estimated by the Data Review Board by estimating conditional power after 15 C-DM and 10 C-HT subjects had completed the study. However, the pattern of subject enrollments into the study is such that the study will be fully enrolled (50 subjects) prior to reaching the predefined point of sample size re-estimation. Therefore, the Data Review Board will not re-estimate the sample size.

7.4 Data Handling/Imputation Methods

7.4.1 Missing Data on Primary Endpoint Measures

$AUC_{glucose}$

$AUC_{glucose}$ measures will be obtained for all subjects (C-DM and C-HT) at Baseline and Weeks 6, 10, 16, 24 (or early termination visit).

When calculating the AUC at a given time point, the following rules for handling missing data will be applied:

. If data for fasting plasma glucose (time point 0) or the plasma glucose 30 minute post oral glucose administration time point are missing, no AUC calculation will be performed for that particular visit and the AUC for that visit will be counted as missing.

. If glucose concentration data are not available for more than one oGTT (post oral glucose administration) time point (i.e., 30, 60, 90 and 120 minutes) for a particular visit, then the $AUC_{glucose}$ will not be calculated for that visit.

. If data for only one plasma glucose (post oral glucose administration) time point other than the 30 minute time point (i.e., either 60 or 90 minutes) is missing, then AUC will be calculated using available data. For example, if 90 minute value is missing, then a larger trapezoid will be constructed to connect the 60 min time point to the 120 min time point.

. If only the 120 minute plasma glucose is missing, the available time points will be used to calculate $AUC_{glucose}$. In such a case, the 120 minute plasma glucose for the baseline oGTT will be disregarded. $AUC_{glucose}$ will be calculated using time points 0, 30, 60, 90 minutes; $AUC_{glucose\ 0\text{-}90}$ will be used for both baseline and final observation.

When evaluating the primary endpoint among C-DM subjects, an endpoint analysis will be calculated using the change from baseline $AUC_{glucose}$ to the last value for $AUC_{glucose}$ obtained (week 24 or early termination visit), provided that

that last observation occurred no more than 14 days after the patient stopped taking the medication. The 14 day period of time has been chosen because, based on the half-life of mifepristone, 14 days exceeds the expected duration of clearance of the drug from the circulation in the vast majority of people who take the drug. Thus the 14 day limit has been chosen to best reflect the effect of the drug on $AUC_{glucose}$ in situations where the interval between discontinuation of study drug and the last oGTT is prolonged. In the case where the last observation occurred more than 14 days after the patient stopped taking the medication, the most recent prior value of $AUC_{glucose}$ will be used.

7.4.2 Blood Pressure

Blood pressure evaluations will be taken for all subjects (C-DM and C-HT) at screening, baseline, days 7, 14, and 28, and weeks 6,8,10,12,16, 20, and 24 (or early termination).

For the primary endpoint evaluation ofC-HT subjects, an endpoint analysis will be calculated using the change from baseline diastolic blood pressure to the last valid value of diastolic blood pressure obtained (week 24 or early termination visit), provided the last observation occurred no more than 14 days after the patient stopped taking the medication. In the case where the last observation occurred more than 14 days after the patient stopped taking the medication, the most recent prior value of diastolic blood pressure will be used.

…

7.6 Examination of Subgroups

Because of the sample size, no formal statistical analyses will be conducted for subgroups (e.g., treatment effect comparison by age, sex, or race). However, summary tables for primary and key secondary endpoints by sex and age will be provided in the CSR.

…

9.1 Analysis of Primary Endpoints

9.1. Subjects with Cushing's Syndrome co-morbid with Diabetes Mellitus Type 2 or impaired glucose tolerance (C-DM)

The primary endpoint for this subject population will be the change in area under the cure for glucose ($AUC_{glucose}$) from 2-hour oral glucose tolerance tests (oGTT) from baseline to Week 24 in subjects with diabetes/impaired glucose tolerance with or without hypertension at screening. A responder analysis using the efficacy population (mITT) will be used to measure success on this primary efficacy endpoint. A responder will be defined as a subject who experiences at least a 25% decrease in $AUC_{glucose}$ from baseline to week 24. The null and alternative hypotheses are as follows:

Ho: $\pi_{\text{25\% reduction in the glucose AUC at 24 weeks}} \leq 0.2$

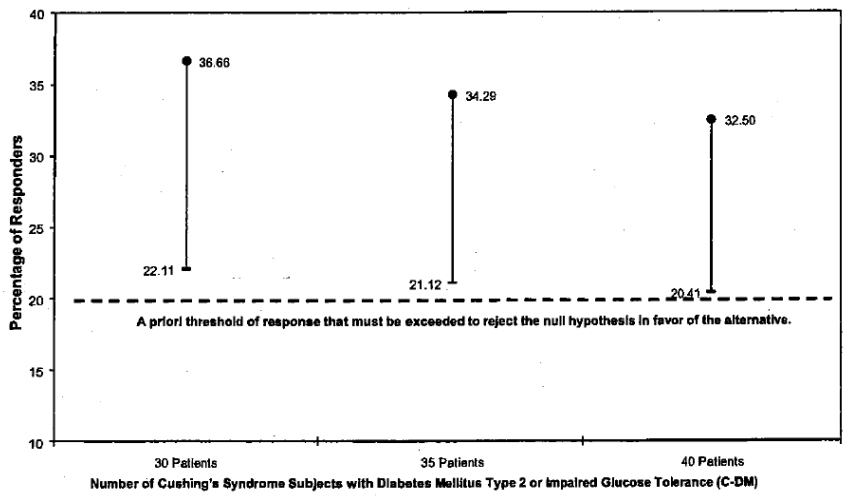
Ha: $\pi_{\text{25\% reduction in the glucose AUC at 24 weeks}} > 0.2$

This null hypothesis will be rejected in favor of the alternative if the lower limit of the exact one-sided 95% binomial confidence interval for the responder rate is greater than 20%.

Although there are anecdotal reports of spontaneous remissions in Cushing's syndrome, the cases are extremely rare. Twenty percent (20%) is an appropriate threshold to test against in this population, given that the spontaneous remission rate in individuals who are eligible for this study is close to 0%. These rare cases often occur in the setting of de novo Cushing's disease and have been largely due to apoplexy (pituitary hemorrhage). Subjects who are enrolled with prior pituitary radiotherapy could theoretically lead to a higher remission rate. Although the criteria used for establishing remission are not standardized, recent reviews document that the usual time to remission in those patients who respond is approximately 2 years depending on the modality of radiotherapy used. While control of hypercortisolism may occur in as many as 50-60% of patients in 3-5 years, responses are gradual, variable and may be delayed for many years. These rates are applicable to specialized centers and overall response rates are likely lower. Because the use of pituitary radiation is not widespread and it is not used in non-pituitary Cushing's Syndrome, only a few subjects enrolled into the study are expected to have received pituitary radiation prior to enrollment. Of those, only a fraction would be expected to achieve a response from radiation. Thus, while possible, an enrolled subject with active Cushing's disease and previous radiation therapy is unlikely to have a remission during the 24 week treatment period of the study. Measurements of ACTH and cortisol production at frequent intervals during the study and at the 6-week follow up safety evaluation will provide an assessment of changes in underlying disease status; pituitary MRIs will provide information regarding pituitary hemorrhage that may have occurred during the study. Based upon these considerations, a lower bound of 20% for the 95% binomial confidence interval was chosen to provide sufficient margin between the rate of spontaneous remissions not due to study drug and the estimate of responder rate due to study drug.

Although a two-sided 95% confidence is appropriate in defining a parameter where both the upper and lower bounds are of substantive interest, this does not reflect the single-arm study design and evaluation criteria for clinical response from the currently ongoing study C1073-400. The spontaneous remission rate in this syndrome is extremely low (<5%), and the a priori threshold of 20% is synonymous with a clinically-meaningful effect. Based on the unidirectional hypothesis specified for this study in this patient population, the patient-based response would need to be robust in order to exceed 20% with 95% confidence. Figure No.1 provides a visual examination of the magnitude of effect required to reject the nun hypothesis.

Figure No.1 An Open Label Study of the Efficacy and Safety of CORLUX®(mifepristone) in the Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome.
Figure No.1: Percentage of C-DM Patients with Exact 95% Binomial Confidence Limits



Prepared: February 17, 2010 (version no. 1.0)

The table presented below contains the point estimate of response required to reject the null hypothesis in favor of the alternative.

LOWER 1-SIDED 95% EXACT BINOMIAL CONFIDENCE INTERVALS EVALUATED BASED ON THE A PRIORI THRESHOLD OF 20%

NUMBER OF SUBJECTS	NUMBER (%) OF SUBJECTS WHO MET THE ENDPOINT	LOWER 1-SIDED 95% EXACT BINOMIAL CONFIDENCE INTERVAL	REJECT THE NULL HYPOTHESIS IN FAVOR OF THE ALTERNATIVE HYPOTHESIS (SUCCESSFUL OUTCOME)
30	11 (36.7%)	22.11%	YES
35	12 (34.3%)	21.12%	YES
40	13 (32.5%)	20.41%	YES

$AUC_{glucose}$ values will be calculated using the plasma glucose concentration data and will be expressed in mg/dL X 2hr. These calculations will be performed programmatically using SAS®softare, version 9.2. The following calculation for the linear trapezoidal AUC model will be used: AUC(0-t) =

$$\sum_{i=1}^{t}\left(\frac{C_i + C_{i-1}}{2}\right)(t_i - t_{i-1})$$, where

t = pre-specified time point C = concentration at time t_i

These resulting $AUC_{glucose}$ values will be verified, using the following detailed formula:

AUCglucose = .5 *(C1 + C2) * (t2 - tl) + .5 * (C2 + C3) * (13 - t2) + .5 * (C3 + C4) * (t4 - 13) + .5 * (C4 + C5) * (t5 - t4).

Where t = pre-specified time point and C = concentration at time t_{i-1}).

Plasma glucose concentration data will be summarized by visit and assessment time and presented in a table. Individual plasma glucose concentrations for each subject as well as peak glucose after oGTT and 2 hour value will be presented in

a listing. Deviations from scheduled sample collections (i.e., actual sample time relative to the glucose administration time versus the nominal collection time) will be calculated and included in a listing. Plots of the mean, median and individual plasma glucose concentrations versus time will be generated by visit for subjects with diabetes/impaired glucose tolerance at screening.

For total AUC calculation, if the AUC is not calculable at week 24, the most recent previous calculable AUC will be used for the primary endpoint in subjects with diabetes mellitus/impaired glucose tolerance. If more than one value is missing at week 24, week 16 will be used as the efficacy measure for the primary endpoint (endpoint analysis). The treatment of missing values is detailed further in Section 7.4. Individual subject $AUC_{glucose}$ values will be listed, and a descriptive summary, including change from baseline for all visits, will be provided by visit in tabular form. Counts of responders versus non-responders and the results of the binomial test will be presented in a table.

In addition, the response characteristics of the change in $AUC_{glucose}$ from 2-hour oral glucose tolerance tests (oGTT) from baseline to Week 24 (or early termination) in C-DM subjects will be presented as cumulative distribution functions (CDF), where change in $AUC_{glucose}$ is calculated by

$$\text{Change in AUC} = \frac{\text{AUC at Week 24 (or early termination) - AUC at Baseline}}{\text{AUC at Baseline}} \times 100\%$$

A decrease in $AUC_{glucose}$ represents an improvement in function; therefore negative values in changes from baseline in $AUC_{glucose}$ represent improvement. The CDF will be presented in a table and a figure. The figure will have the x-axis as the range of percent change in $AUC_{glucose}$ from baseline to Week 24 (or early termination) from the most negative to most positive. The y-axis will show the proportion of subjects with a percent change from baseline of the x value or less. Lines connecting the points will be shown. The table and figure will be produced using the mITT population with imputation methods described in section 7.4.1 and then with observed values in the completer population.

9.1.2 Subjects with Cushing's Syndrome and Co-morbid Hypertension and No Diabetes Mellitus/Impaired Glucose Tolerance

Response on the primary efficacy endpoint for C-HT subjects is defined as either a reduction in diastolic blood pressure. Reduction in Diastolic Blood Pressure Changes in diastolic blood pressure will be analyzed for subjects with hypertension only at screening (i.e., without impaired glucose tolerance or diabetes) as a primary endpoint. A responder analysis of the reduction in diastolic blood pressure from baseline to week 24 will be performed for the efficacy population. A responder will be defied as a subject who experiences a 5 mmHg or greater decline in diastolic blood pressure from baseline to week 24.

$$H_o: \pi_{\text{5 mmHg reduction in the diastolic blood pressure at 24 weeks}} \leq 0.2$$

$$H_a: \pi_{\text{5 mmHg reduction in the diastolic blood pressure at 24 weeks}} > 0.2$$

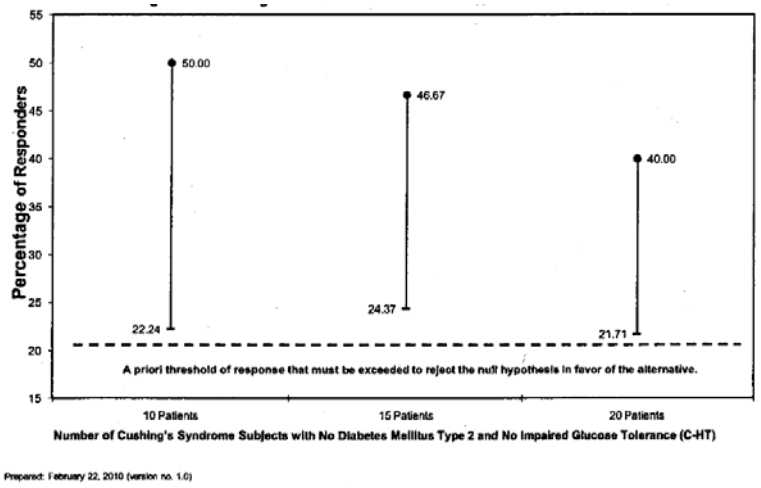
This null hypothesis will be rejected in favor of the alternative if the lower limit of the exact one-sided 95% binomial confidence interval for the responder rate is greater 20%. The lower limit of20% is an appropriate threshold to test against in

this population because it exceeds a conservative estimate of the spontaneous remission rate expected in subjects participating in this study. Figure No.2 provides a visual examination of the magnitude of effect required to reject the null hypothesis.

Figure No.2

An Open Label Study of the Efficacy and Safety of CORLUX® (mifepristone) in the Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome.

Fig No.2: Percentage of C.HT Patients with Exact 95% Binomial Confidence Limits



The table presented below contains the point estimate of response required to reject the null hypothesis in favor of the alternative.

LOWER 1-SIDED 95% EXACT BINOMIAL CONFIDENCE INTERVALS EVALUATED BASED ON THE A PRIORI THRESHOLD OF 20%

NUMBER OF SUBJECTS	NUMBER (%) OF SUBJECTS WHO MET THE ENDPOINT	LOWER 1-SIDED 95% EXACT BINOMIAL CONFIDENCE INTERVAL	REJECT THE NULL HYPOTHESIS IN FAVOR OF THE ALTERNATIVE HYPOTHESIS (SUCCESSFUL OUTCOME)
10	5 (50.0%)	22.24%	YES
15	7 (46.7%)	24.37%	YES
20	8 (40.0%)	21.71%	YES

For subjects treated with spironolactone to treat hypokalemia within the 24 week evaluation period, the effect of spironolactone on blood pressure will be estimated. Subjects with and without exposure to spironolactone will be summarized in a table.

Diastolic blood pressure data will be summarized by visit and change from baseline for both diabetic/impaired glucose tolerant subjects and for hypertensive subjects. These data will be presented in tables for each group. Systolic blood pressure data will be handled in the same manner with results presented in tab1es. Individual blood pressure results for each subject will be presented in a listing.

In addition, the response characteristics of the changes in diastolic blood pressure from baseline to Week 24 (or early termination) in C-HT subjects will be presented as cumulative distribution functions (CDF), where change in diastolic blood pressure is calculated by

Change in DBP = DBP at Week 24 (or early termination) - DBP at Baseline.

A decrease in diastolic blood pressure represents an improvement in function; therefore negative values in changes from baseline represent improvement. The CDF will be presented in a table and a figure. The figure will have the x-axis as the range of change in diastolic blood pressure from baseline to Week 24 or early termination from the most negative to most positive. The y-axis will show the proportion of subjects with a change from baseline of the x value or less. Lines connecting the points will be shown. The table and figure will be produced using the mITT population with imputation methods described in section 7.4.2 and then with observed values in the completer population.

9.2 Analysis of Secondary Endpoints

9.2.1 Key Secondary Endpoint: Clinical Improvement

The key secondary assessment of efficacy will be conducted by the Data Review Board, an independent 3-member panel of expert reviewers with expertise in Cushing's Syndrome. The reviewers will be blinded to the dates and sequence of all visits except the baseline and 6-week safety follow-up visits (that occur 6 weeks after discontinuation of study drug). The reviewers will use ordinal level ratings (-1 through + 1) to assess the subject outcome at defined visits. Each reviewer will assess the data available for the baseline, week 6, week 10, week 16, week 24/early termination, and 6-week follow-up visits, and assign an overall score for each visit as follows: -1 = worse than baseline; 0 = unchanged from baseline; + i = clinically significant improvement. These assessments will be based on data subject efficacy profiles presented to the Data Review Board.

The median of the three scores will be calculated, and the subject will have demonstrated clinical improvement if the median score is +1. Equivalently, this means that the subject will have demonstrated clinical improvement for this study if at least two of the three reviewers rate the subject as + 1 (clinically significant improvement).

A responder analysis will be used to determine the success of this secondary endpoint. For this secondary efficacy endpoint, a responder will be defined as a subject whose median reviewer score is + 1 at any reviewed visit after baseline through week 24/early termination. This efficacy measurement will be declared successful if the lower limit of the exact 95% binomial confidence interval for the responder rate is greater than or equal to 30%. The 30% level was chosen because assessments of change in clinical status made by the Data Review Board may result in a more variability than that resulting from changes in the primary efficacy variable where the lower level of the confidence interval was set at 20%."

Changes to the Planned Analyses

The following changes were made to the planned analyses:

· Although the protocol stated that assays may be performed to measure concomitant medication levels in subjects taking amlodipine, hydrocodone, ibuprofen, omeprazole, and/or rosuvastatin (to be performed on aliquots of the samples obtained for mifepristone trough levels), these assays have not been conducted.

· Exploratory regression analyses were planned to describe change over time for the following variables: weight, HbAlc, $AUC_{glucose}$, and diastolic and systolic blood pressure.

In cases where the data did not meet assumptions necessary for this analysis (e.g., linearity, homogeneity of variance), either 1) the raw data were log-transformed to meet the assumptions (population regressions on weight and AUCgJucose) or 2) scatter plots were substituted for the regression analysis (individual subject regressions on all five variables).

. One table (symptoms that could be associated with adrenal insufficiency requiring glucocorticoid treatment) and one listing (protocol deviations) were generated by Corcept rather than by (b) (4)

· The table for extent of study drug exposure was simplified.

. Tables summarizing subjects taking spironolactone for treatment of hypokalemia were changed so that only the 24-week measurement was presented and eplerenone was added to the analysis of effect.

. Adverse events by SOC and PT had the time frame for assessing non-serious AEs expanded from 2 days after end of treatment to 14 days after end of treatment.

Results and Conclusions

Primary Efficacy Analyses

Subjects with Diabetes Mellitus and/or Impaired Glucose Tolerance (**C-DM Cohort**)

The primary endpoint for subjects with Cushing's syndrome and diabetes or impaired glucose tolerance at screening (C-DM cohort) was the change in AUC for glucose ($AUC_{glucose}$) from baseline to Week 24/ET for the 2-hour oGTT in the mITT population.

For the C-DM cohort (mITT), the number and percent of responders (those who had a 25% or more decrease from baseline in $AUC_{glucose}$ at Week 24/ET) are summarized in Table below. Fifteen subjects (60.0%) achieved this endpoint (Subjects 03-004, 06-003, 07-003, 08-011, 08-013,09-001,10-001,10-002,10-004,11-002,11-003, 17-002,18-001 23-001, and 24-006). Sponsor’s presentations follow after this reviewer’s comments and presentations.

Reviewer's Analyses and Comments: We advised the sponsor to apply 95% 2-sided or 97.5% 1-sided confidence intervals. The sponsor neglected our advice; so, I have computed the same by which the lower confidence interval is 40.4%, still, supporting the sponsor's claim.

AUC-GLUCOSE

Variable	n	Mean	Standard Error	Lower CI	Upper CI	T value	Pr>|T|
Baseline	23	30670	1885.1	26975.63	34365.23		
Change from baseline	23	-8722.2	2276.6	-13184.33	-4260.02	-3.83	.0009
Percent Change from baseline	23	-25.0	5.54	-35.8473	-14.11	-4.51	.0002

The confidence intervals above do not include zero, showing statistical significance of the mean change and mean percent change from baseline in AUC-Glucose.

HbA1c

Variable	n	Mean	Standard Error	Lower CI	Upper CI	T value	Pr>|T|
Baseline	24	7.36	.2977	6.776	7.944		
Change from baseline	24	-1.11	.2326	-1.564	-.652	-4.76	.0001
Percent Change from baseline	24	-18.23	3.9676	-26.011	-10.458	-4.60	.0001

The confidence intervals above do not include zero, showing statistical significance of the mean change and mean percent change from baseline in HbA1c.

Sponsor's Presentations

A response in $AUC_{glucose}$ was observed in 60% of the subjects (95% CI lower bound, 42%).

Because the lower bound of the 95% CI was greater than 20%, this response rate of 60% was statistically significant. The results of the responder analysis for the ITT and Completer

populations were similar to those for the mITT population and were also statistically significant (52% response rate and 35% lower bound 95% CI for the ITT population, Table 14.2.3.3, and 65% response rate and 44% lower bound 95% CI for the Completer population, Table 14.2.3.2).

Responder Analysis: Subjects Who Had at Least a 25% Decrease from Baseline in $AUC_{glucose}$ at Week 24/ET (mITT Population)

	C-DM N=25 n (%)	Lower Bound 1-sided 95% Exact Binomial CI
Subjects who had at least a 25% decrease from baseline in $AUC_{glucose}$ at Week 24/ET		
Responder	15 (60.0)	41.68%
Non-responder	10 (40.0)	

AUC = area under the concentration-time curve; C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); CI = confidence interval; ET = early termination; mITT = modified intent-to-treat; oGTT = oral glucose tolerance test.

A responder was defined as a subject who had a least a 25% decrease in $AUC_{glucose}$ at Week 24/ET compared with baseline. The null hypothesis was to be rejected in favor of the alternative if the lower limit of the exact 1-sided 95% binomial CI for the responder rate was > 20%.

Week 24/ET values include imputed data and are used for the efficacy analysis.

Source: Table 14.2.3.1. Supporting data (a summary of the 75-gram oGTT test data and $AUC_{glucose}$ values by visit) are presented for the mITT Population in Table 14.2.1 and plasma glucose concentration data from the 2-hour oGTT are listed for all subjects in the DM cohort in Listing 16.2.6.1.1.

The overall reduction in $AUC_{glucose}$ in the mITT C-DM cohort (actual values and percent reduction) is shown in Table below. At baseline, the median $AUC_{glucose}$ value was 30330.0 mg/dL (over 2 hours), which decreased to 23655.0 mg/dL at Week 6 and to 19950.0 mg/dL at Week 16. At Week 24/ET, the median $AUC_{glucose}$ value was 20655.0 mg/dL, which represented a 36% reduction over the course of the study. Many subjects had large responses, with reductions of up to 40 to 60% or more (Table 14.2.3.4).

Decreases tended to be rapid for most subjects. There were similar reductions in $AUC_{glucose}$ in the C-DM cohorts of the ITT and Completer populations (Tables 14.2.3.6 and 14.2.3.5, respectively). Figures 14.2.1.1 and 14.2.1.2 show plots of the mean oGTT glucose values in C-DM cohort of the mITT and ITT populations, respectively.

Individual subject data for percent reduction in $AUC_{glucose}$ are summarized in Table 14.2.3.4 (mITT Population), Table 14.2.3.6 (ITT Population), and Table 14.2.3.5 (Completer Population). Figure 14.2.1.5 shows plots of oGTT glucose values for individual C-DM subjects. A listing of all $AUC_{glucose}$ data can be found in Listing 16.2.6.1.2. Listings of plasma glucose concentration data from 2 hour oGTT and deviations from the plasma glucose sampling times can be found in Listing 16.2.6.1.1 and Listing 16.2.6.1.3, respectively.

Percent Reduction in $AUC_{glucose}$ in C-DM Subjects by Visit (mITT Population)

Visit	$AUC_{glucose}$ (mg/dL over 2 hours)	% Reduction from Baseline
Baseline		
n	25	
Mean (SD)	32185.2 (10191.74)	
Median	30330.0	
Min, Max	18480, 57795	
Week 6		
n	23	23
Mean (SD)	24703.0 (8823.31)	-21.776 (19.0642)
Median	23655.0	-19.469
Min, Max	13890, 52965	-50.69, 14.23
Week 10		
n	20	20
Mean (SD)	23384.3 (8434.72)	-23.576 (22.6960)
Median	22327.5	-29.004
Min, Max	13710, 43500	-52.06, 35.58
Week 16		
n	20	20
Mean (SD)	21625.5 (7398.09)	-29.186 (23.6989)
Median	19950.0	-31.099
Min, Max	12855, 44100	-61.82, 37.45
Week 24/ET		
n	24	24
Mean (SD)	22365.0 (7757.35)	-27.038 (26.2438)
Median	20655.0	-36.129
Min, Max	11475, 42750	-68.67, 33.24

AUC = area under the concentration-time curve; C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); ET = early termination; Max = maximum; Min = minimum; mITT = modified intent-to-treat; SD = standard deviation.

Percent reduction = ([$AUC_{glucose}$ at Week XX - $AUC_{glucose}$ at baseline] / [$AUC_{glucose}$ at baseline]) × 100.

Negative number indicates improvement.

Week 24/ET values include imputed data and are used for the efficacy analysis.

Source: Table 14.2.3.4.

§

The cumulative distribution for percent reduction in $AUC_{glucose}$ at Week 24/ET is presented in Table and in Figure below. The data in this table are sorted by most improvement to least improvement. A large percentage (>80%) of subjects in both mITT and ITT populations had some degree of improvement (reduction) in $AUC_{glucose}$ (21 of 24 subjects for mITT and 22 of 27 subjects for ITT). The range of percent improvement (reduction) was 0.8% to 69% for both the mITT and ITT populations. Among subjects who had an improvement in $AUC_{glucose}$ greater than or equal to the predefined threshold of 25%, the median improvement in $AUC_{glucose}$ was 43.9%. Seventeen (71 %) of 24 subjects in the mITT population and 17 (63%) of 27 subjects in the ITT population had a reduction in $AUC_{glucose}$ that exceeded 15%.

Cumulative Distribution Function for Percent Reduction in $AUC_{glucose}$ at Week 24/ET in C-DM Subjects (mITT and ITT Populations)

mITT Population			ITT Population		
% Reduction from Baseline	Cumulative Distribution of Change n (%)	Improved/ Worsened	% Reduction from Baseline	Cumulative Distribution of Change n (%)	Improved/ Worsened
-68.7	1 (4.17)	Improved	-68.7	1 (3.70)	Improved
-60.6	2 (8.33)	Improved	-60.6	2 (7.41)	Improved
-50.7	3 (12.50)	Improved	-50.7	3 (11.11)	Improved
-48.2	4 (16.67)	Improved	-48.2	4 (14.81)	Improved
-47.0	5 (20.83)	Improved	-47.0	5 (18.52)	Improved
-46.5	6 (25.00)	Improved	-46.5	6 (22.22)	Improved
-45.3	7 (29.17)	Improved	-45.3	7 (25.93)	Improved
-43.9	8 (33.33)	Improved	-43.9	8 (29.63)	Improved
-42.4	9 (37.50)	Improved	-42.4	9 (33.33)	Improved
-41.2	10 (41.67)	Improved	-41.2	10 (37.04)	Improved
-37.1	11 (45.83)	Improved	-37.1	11 (40.74)	Improved
-36.9	12 (50.00)	Improved	-36.9	12 (44.44)	Improved
-35.3	13 (54.17)	Improved	-35.3	13 (48.15)	Improved
-28.0	14 (58.33)	Improved	-28.0	14 (51.85)	Improved
-25.0	15 (62.50)	Improved	-25.0	15 (55.56)	Improved
-24.2	16 (66.67)	Improved	-24.2	16 (59.26)	Improved
-19.8	17 (70.83)	Improved	-19.8	17 (62.96)	Improved
-4.4	18 (75.00)	Improved	-8.3	18 (66.67)	Improved
-3.8	19 (79.17)	Improved	-4.4	19 (70.37)	Improved
-3.5	20 (83.33)	Improved	-3.8	20 (74.07)	Improved
-0.8	21 (87.50)	Improved	-3.5	21 (77.78)	Improved
4.7	22 (91.67)	Worsened	-0.8	22 (81.48)	Improved
26.5	23 (95.83)	Worsened	2.0	23 (85.19)	Worsened
33.2	24 (100.00)	Worsened	4.7	24 (88.89)	Worsened
			6.8	25 (92.59)	Worsened
			26.5	26 (96.30)	Worsened
			33.2	27 (100.0)	Worsened

AUC = area under the concentration-time curve; C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); ET = early termination; ITT = intent-to-treat; mITT = modified intent-to-treat.

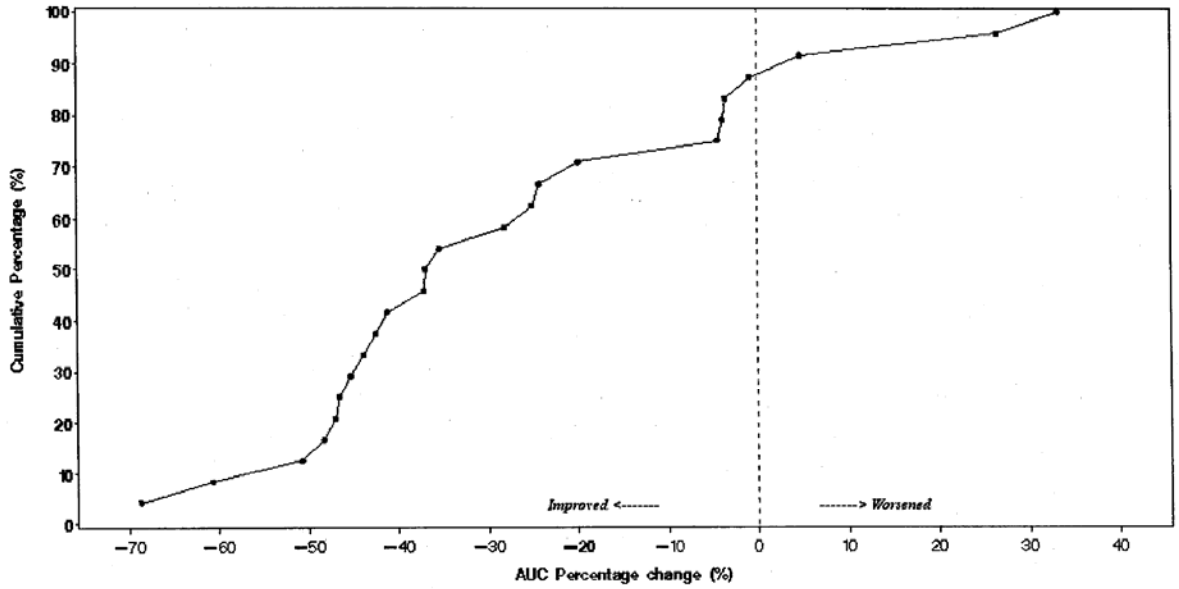
Percent reduction = ([$AUC_{glucose}$ at Week XX - $AUC_{glucose}$ at baseline] / [$AUC_{glucose}$ at baseline]) × 100. A negative value in percent reduction $AUC_{glucose}$ denotes an improvement in function.

Cumulative distribution column displays the number of subjects (%) with a percent reduction in $AUC_{glucose} \leq x\%$, where x% is the value listed in the first column.

Two subjects (08-003, 20-002) were excluded from the ITT analysis and one subject (08-003) was excluded from the mITT analysis because they did not have AUC values post baseline.

Source: Table 14.2.3.7 and Table 14.2.3.9.

Figure 1. Cumulative Distribution Function for Percent Change in $AUC_{glucose}$ from Baseline to Week 24/ET: C-DM Cohort (mITT Population)



Source: Table 14.2.3.7.

Subjects with Hypertension (C-HT Cohort)

The primary efficacy variable in subjects with Cushing's syndrome and a diagnosis of hypertension only (without diabetes or impaired glucose tolerance), i.e., the C-HT cohort, was reduction in diastolic blood pressure.

For the C-HT cohort (mITT), the number and percent of responders (those who had at least a 5 mmHg reduction in diastolic blood pressure at Week 24/ET) are summarized in Table 16. Eight subjects (38.1 %) achieved this endpoint (Subjects 03-002, 06-002, 07-010, 08-004, 11- 004, 16-002, 22-001, and 24-002). Subject 10-003 demonstrated a decrease of > 5 mmHg in diastolic blood pressure when the blood pressure readings from Day 1 were compared to Week 24. However, the subject started a new antihypertensive medication at Week 23, followed by the diastolic blood pressure decrease at Week 24. This subject was counted as a non-responder.

Sponsor's presentations follow after this reviewer's comments.

Reviewer's Analyses and Comments: We advised the sponsor to apply 95% 2-sided or 97.5% 1-sided confidence intervals. The sponsor neglected our advice; so, I have computed the same by which the lower confidence interval is 16.8%, nullifying the sponsor's claim. The 95% confidence interval does not include the 20% margin. The sponsor's presentations for C-HT patients may not even be read.

Diastolic Blood Pressure

Variable	n	Mean	Standard Error	Lower CI	Upper CI	T value	Pr>|T|
Baseline	21	82.86	2.49	77.97	87.73	33.259	
Change from baseline	21	-0.05	2.34	-4.64	4.55	-0.02	.984
Percent Change from baseline	21	-0.47	2.81	-5.98	5.04	0.17	.869

Above Table shows that diastolic blood pressure did not change from baseline.

Sponsor's Presentations

A response for diastolic blood pressure was observed in 38% of the subjects (95% CI lower bound, 21 %). Because the lower bound 95% CI was greater than 20%, this response rate of 38% was statistically significant. The results of the responder analysis for the ITT and Completer populations were supportive of those for the mITT population and were also statistically significant (38% response rate and 21 % lower bound CI for the ITT population, Table 14.2.5.3, and 54% response rate and 29% lower bound 95% CI for the Completer population, Table 14.2.5.2). Note: the ITT and the mITT populations were the same for the C-HT cohort.

Responder Analysis: Subjects Who Had at Least a 5 mmHg Reduction in Diastolic Blood Pressure at Week 24/ET (mITT Population)

	C-HT N=21 n (%)	**Lower Bound 1-sided 95% Exact Binomial CI**
Subjects who had at least a 5 mmHg reduction from baseline in diastolic blood pressure at Week 24/ET		
Responder	8 (38.1)	20.57%
Non-responder	13 (61.9)	

C-HT = subjects with Cushing's syndrome and diagnosis of hypertension; CI = confidence interval; ET = early termination; mITT = modified intent-to-treat

A responder was defined as a subject who had a least a 5 mmHg reduction in diastolic blood pressure at Week 24/ET compared with baseline. The null hypothesis was to be rejected in favor of the alternative if the lower limit of the exact 1-sided 95% binomial CI for the responder rate was > 20%.

Week 24/ET values include imputed data and are used for the efficacy analysis.

Source: Table 14.2.5.1.

The overall change in diastolic blood pressure in the C-HT cohort (actual values and change from baseline) is shown in Table 17. At baseline, the mean value (± SD) was 82.9 ± 11.42 mmHg. A mean decrease was not observed until Week 16, at which time the mean value was 81.6 ± 12.48 mmHg. At Week 24/ET, the mean diastolic blood pressure value was 82.8 ± 13.16

mmHg, which did not represent any change from baseline in the cohort as a whole. The range of diastolic blood pressure was wide at baseline and persisted throughout the study; median diastolic blood pressure decreased from 87 mmHg at baseline to 81 mmHg at Week 24/ET. The results of the diastolic blood pressure analyses were similar for the Completer population (Tables 14.2.4.2 and 14.2.5.5) as for the mITT and ITT populations.

Summary of Diastolic Blood Pressure in C-HT Subjects by Visit (mITT Population)

Visit	C-HT (N=21)	Change from Baseline
Baseline		
n	21	
Mean (SD)	82.9 (11.42)	
Median	87	
Min, Max	62, 108	
Week 6		
n	20	20
Mean (SD)	82.9 (12.80)	-0.8 (10.12)
Median	86.5	0.5
Min, Max	47, 104	-16, 17
Week 10		
n	18	18
Mean (SD)	86.1 (11.50)	1 (11.34)
Median	88.5	2.5
Min, Max	64, 103	-20, 21
Week 16		
n	16	16
Mean (SD)	81.6 (12.48)	-3.2 (12.28)
Median	81	-4
Min, Max	56, 102	-28, 14
Week 24/ET		
n	21	21
Mean (SD)	82.8 (13.16)	0 (10.74)
Median	81	-4
Min, Max	61, 108	-20, 20

C-HT = subjects with Cushing's syndrome and hypertension; ET = Early termination; Max = maximum; Min = minimum; mITT = modified intent-to-treat; SD = standard deviation.

Week 24/ET values include imputed data and are used for the efficacy analysis.

Source: Table 14.2.4.1.

The cumulative distribution results for the C-HT cohort are presented in Table below for the mITT and ITT populations. The results were similar for the Completer population (Table 14.2.5.5).

Change from Baseline (mmHg)	Cumulative Distribution of Change n (%)	Improved/Worsened
-20.0	1 (4.76)	Improved
-19.0	2 (9.52)	Improved
-9.0	3 (14.29)	Improved
-8.0	4 (19.05)	Improved
-7.0	5 (23.81)	Improved
-6.0	7 (33.33)	Improved
-5.0	9 (42.86)	Improved
-4.0	11 (52.38)	Improved
0.0	12 (57.14)	No change
3.0	13 (61.90)	Worsened
5.0	14 (66.67)	Worsened
6.0	15 (71.43)	Worsened
8.0	16 (76.19)	Worsened
9.0	17 (80.95)	Worsened
10.0	18 (85.71)	Worsened
14.0	19 (90.48)	Worsened
17.0	20 (95.24)	Worsened
20.0	21 (100.00)	Worsened

C-HT = subjects with Cushing's syndrome and hypertension; ET = early termination; mITT = modified intent-to-treat; ITT = intent-to treat.

Reduction = diastolic blood pressure at Week 24/ET - diastolic blood pressure at baseline.

Cumulative distribution column displays the number of subjects (%) with a reduction in diastolic blood pressure ≤ 5 mmHg, where x% is the value listed in the first column.

Week 24/ET values include imputed data and are used for the efficacy analyses.

Source: Table 14.2.5.4 and Table 14.2.5.6

A summary of diastolic blood pressure at baseline and Week 24/ET for subjects with and without spironolactone treatment is presented in Table 14.2.6.1 (mITT population), Table 14.2.6.3 (ITT population), and Table 14.2.6.2 (Completer population). Because hypokalemia is a known side effect of treatment with mifepristone, spironolactone use was permitted by study protocol to treat hypokalemia not responsive to potassium supplementation alone. Four C-HT subjects received spironolactone contemporaneously with efficacy blood pressure readings, one of whom had been on spironolactone prior to study start and remained on this medication at a stable dose throughout the study (Subject07-006). Two subjects (07-010 and 11-004) who had a reduction in diastolic blood pressure and met the responder criteria for this endpoint received spironolactone for treatment of hypokalemia (Listing 16.2.6.2.1 and Listing 16.2.4.4.2). Subject 07-010 had a >=5 mmHg reduction in diastolic blood pressure compared to baseline starting at the Day 7 visit (21 April 2010) that continued through Week 24 (29 September 2010). Spironolactone treatment for hypokalemia/edema was initiated on 7 May 2010 and was ongoing at the end of the study. Subject 11-004 had a reduction in lisinopril dosage (for the treatment of hypertension) from 40 mg QD to 10 mg QD at Week 20 and a ~ 5 mmHg reduction in diastolic blood pressure compared to baseline at Week 24 (11 November 2010). Spironolactone treatment was initiated on 9 July 2010 (after the Week 6 visit) for the treatment of edema and discontinued on 19 November 2010. The subject's diastolic blood pressure was increased at Weeks 8, 10, 12, 16, and 20, but was decreased at Week 24.

Key Secondary Efficacy Analysis

Clinical Improvement as Assessed by the Data Review Board

A responder analysis was used to determine clinical improvement from the assessments of the DRB comprised of three independent experts on Cushing's syndrome. The DRB performed a review of eight categories of clinical parameters to evaluate whether a subject's signs and symptoms of Cushing's syndrome had changed. For the 33 subjects who consented to have photographs taken, their photographs were also reviewed by the DRB. The categories of assessment were:

I. Assessment of glucose homeostasis

2. Assessment of blood pressure

3. Assessment of lipids

4. Change in weight and body composition

5. Clinical scoring and appearance (eg, acne, hirsutism, striae, Cushingoid appearance)

6. Strength assessment

7. Psychiatric and quality of life assessment

8. Metabolic bone assessment

A responder was defined as a subject whose median reviewer score was + 1 at any reviewed visit after baseline through Week 24/ET (see Section 9.7.1.6.3). The DRB assessed only those subjects who received a total of at least 30 days of mifepristone during the 24-week treatment period (ie, the mITT population); thus the data from the four subjects not included in the mITT population (15-001, 15-005,20-002, and 24-005; Section 11.1) were not reviewed, and these subjects were considered nonresponders for the DRB analysis.

The number and percent of responders with clinical improvement scores of + 1 are shown in Table 19. In the overall mITT population, the response rate was 87% (lower bound 95% CI, 76%). The percent of responders was 92% in the C-DM cohort (lower bound 95% CI, 77%) and was 81 % in the C-HT cohort (lower bound 95% CI, 62%). Because the lower bound 95% CI was greater than 30% for these analyses (set as the a priori lower bound), these response rates were statistically significant. Results for the ITT and Completer populations were supportive of those for the mITT population, also showing statistically significant responder rates (Table 14.2.8.3 and Table 14.2.8.2, respectively).

Comment: The definition of a Responder in the key secondary efficacy variable above: "A responder was defined as a subject whose median reviewer score was + 1 at any reviewed visit

after baseline through Week 24/ET" with the phrase "at any reviewed visit," gives multiple opportunities for a success and is not as dependable as a response at any one time-point. Therefore, the results of this key variable, if at all are allowed to be in the labeling, this point should be emphasized.

Responder Analysis: Number of Subjects With or Without a Median Clinical Improvement Score of +l at Any Reviewed Visit (mITT Population)

	n (%)	Lower Bound 1-sided 95% Exact Binomial CI
Responders had median clinical improvement score of +1 at any reviewed visit		
Combined cohorts		
Responder	40 (87.0)	75.87%
Non-responder	6 (13.0)	
C-DM cohort		
Responder	23 (92.0)	76.90%
Non-responder	2 (8.0)	
C-HT cohort		
Responder	17 (81.0)	61.56%
Non-responder	4 (19.0)	

C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); C-HT = subjects with Cushing's syndrome and hypertension; CI = confidence interval; mITT = modified intent-to-treat.

The null hypothesis of < 30% of subjects within a cohort having a median clinical improvement score of +1 at any reviewed visit was to be rejected in favor of the alternative if the lower limit of the exact 1-sided 95% binomial CI for the responder rate was ≥ 30%.

Source: Table 14.2.8.1.

The median DRB reviewer scores for clinical improvement by visit for both the C-DM and C-HT cohorts, which contributed to the overall responder analyses, are presented in Table 20. A total of 33 of 46 (72%) of subjects had a median score of + I at Week 24/ET; compared with any other visit, the Week 24/ET visit had the highest number and percentage of improved subjects. Eleven of the subjects had an improvement in DRB score early (at Week 6) that persisted throughout the study; another six subjects had improvement in DRB score at Week 10 that persisted (see Listing 16.2.6.3). Three subjects had a non-sustained improvement with a median score of +1 at a visit prior to Week 24/ET and a median score of 0 at Week 24/ET (Subjects 08-005, 08-014, 17-002). Throughout the study, only one subject (22-003) was rated by the DRB as being worse at Week 24/ET than at baseline.

Median Scores of Data Review Board for Clinical Improvement by Visit (mITT Population)

	C-DM (N=25)[a] n (%)	C-HT (N=21)[a] n (%)	Total (N=46)[a] n (%)
Week 6 (n=44)			
-1 = worse than baseline	0 (0.0)	0 (0.0)	0 (0.0)
0 = unchanged from baseline	12 (48.0)	13 (61.9)	25 (54.3)
+1 = clinically significant improvement from baseline	12 (48.0)	7 (33.3)	19 (41.3)
Data not available	1 (4.0)	1 (4.8)	2 (4.3)
Week 10 (n=38)			
-1 = worse than baseline	0 (0.0)	0 (0.0)	0 (0.0)
0 = unchanged from baseline	10 (40.0)	9 (42.9)	19 (41.3)
+1 = clinically significant improvement from baseline	10 (40.0)	9 (42.9)	19 (41.3)
Data not available	5 (20.0)	3 (14.3)	8 (17.4)
Week 16 (n=36)			
-1 = worse than baseline	0 (0.0)	0 (0.0)	0 (0.0)
0 = unchanged from baseline	3 (12.0)	5 (23.8)	8 (17.4)
+1 = clinically significant improvement from baseline	17 (68.0)	11 (52.4)	28 (60.9)
Data not available	5 (20.0)	5 (23.8)	10 (21.7)
Week 24/ET (n=40)			
-1 = worse than baseline	0 (0.0)	1 (4.8)	1 (2.2)
0 = unchanged from baseline	2 (8.0)	4 (19.0)	6 (13.0)
+1 = clinically significant improvement from baseline	19 (76.0)	14 (66.7)	33 (71.7)
Data not available	4 (16.0)	2 (9.5)	6 (13.0)

C-DM = subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance); C-HT = subjects with Cushing's syndrome and hypertension; ET = early termination; mITT = modified intent-to-treat.

A subject was considered to have demonstrated clinical improvement if at least 2 of the 3 reviewers rated the subjects as +1 (clinically significant improvement).

Week 24/ET values include imputed data and are used for this efficacy analysis.

a All mITT subjects had at least one DRB evaluation. Data are not available for some subjects at some times/visits due to missed visit.

Source: Table 14.2.7 and Listing 16.2.6.3.

3.3 Evaluation of Safety

I do not perform any formal safety evaluation.

3.4 Benefit: Risk Assessment (Optional)

I cannot comment on this at this stage in my individual review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The sponsor stated, "Because of the sample size, no formal statistical analyses will be conducted for subgroups (e.g., treatment effect comparison by age, sex, or race). However, summary tables for primary and key secondary endpoints by sex and age will be provided in the CSR."

Results of the demographic characteristics (at baseline) and other prognostic factors were presented before. There were no significant imbalances between the treatment groups.

4.1 Gender, Race, Age, and Geographic Region

Subgroup analyses were not done and provided for Race and Geographic Region because the sponsor thought that those were not meaningful because of very small sample sizes. The Gender and Age subgroup results provided by the sponsor are presented below. Statistical inference/conclusions about differences in subgroup responses, from such meager data, do not seem to be wise, although there were numerical differences.

Number of Subjects With or Without at least a 25% Decrease from Baseline in AUCglucose at Week 24 (or Early Termination) by Age:
C-DM Cohort
mITT Population

Statistic	C-DM Age <= 46 (N=17)	C-DM Age > 46 (N=8)	C-DM Overall (N=25)
Subjects with or without at least 25% Reduction from Baseline in AUCglucose at Week 24 (or Early Termination)			
Responder n (%)	11 (61.1%)	4 (57.1%)	15 (60.0%)
Non-Responder n (%)	7 (38.9%)	3 (42.9%)	10 (40.0%)

Note: 1. C-DM group includes subjects with diabetes mellitus type 2 and/or impaired glucose tolerance at screening and Day 1 as determined by two or more abnormal oGTTs.
2. Age Group is broken by tertiles. Low and medium = Age Group 1 (Age <= 46 years), High = Age Group 2 (Age > 46years).
Source: Listings 16.2.6.1.2 and 16.2.4.1

Number of Subjects With or Without at least a 25% Decrease from Baseline in AUCglucose at Week 24 (or Early Termination) by Sex:
C-DM Cohort
mITT Population

Statistic	C-DM Male (N=6)	C-DM Female (N=19)	C-DM Overall (N=25)
Subjects with or without a 25% Reduction from Baseline in AUCglucose at Week 24 (or Early Termination)			
Responder n (%)	6 (100.0%)	9 (47.4%)	15 (60.0%)
Non-Responder n (%)	0 (0.0%)	10 (52.6%)	10 (40.0%)

Note: 1. C-DM group includes subjects with diabetes mellitus type 2 and/or impaired glucose tolerance at screening and Day 1 as determined by two or more abnormal oGTTs.
Source: Listings 16.2.6.1.2 and 16.2.4.1

Comment: Numerically, there seems to be differences in response with respect to Gender, with 100% response in males and 47.4% response in females.

Number of Subjects With or Without at least a 5 mmHg Reduction in Diastolic Blood Pressure at Week 24 (or Early Termination) by Age: C-HT Cohort mITT Population

	C-HT		
	Age <= 50 (N=15)	Age > 50 (N=6)	Overall (N=21)
Subjects with or without at least a 5 mmHg Reduction in Diastolic Blood Pressure at Week 24 (or Early Termination)			
Responder n (%)	5 (33.3%)	3 (50.0%)	8 (38.1%)
Non-Responder n (%)	10 (66.7%)	3 (50.0%)	13 (61.9%)

Note: 1. C-HT group includes subjects with diagnosis of hypertension at screening but without diabetes mellitus type 2 and/or impaired glucose tolerance.
2. Age Group is broken by tertiles. Low and medium = Age Group 1 (Age <= 50 years), High = Age Group 2 (Age > 50years).
Source: Listings 16.2.6.2.1 and 16.2.4.1

Number of Subjects With or Without at least a 5 mmHg Reduction in Diastolic Blood Pressure at Week 24 (or Early Termination) by Sex: C-HT Cohort mITT Population

	C-HT		
	Male (N=8)	Female (N=13)	Overall (N=21)
Subjects with or without at least a 5 mmHg Reduction in Diastolic Blood Pressure at Week 24 (or Early Termination)			
Responder n (%)	4 (50.0%)	4 (30.8%)	8 (38.1%)
Non-Responder n (%)	4 (50.0%)	9 (69.2%)	13 (61.9%)

Note: 1. C-HT group includes subjects with diagnosis of hypertension at screening but without diabetes mellitus type 2 and/or impaired glucose tolerance.
2. Age Group is broken by tertiles. Low and medium = Age Group 1 (Age <= 50 years), High = Age Group 2 (Age > 50years).
Source: Listings 16.2.6.2.1 and 16.2.4.1

Number of Subjects With or Without a Median Clinical Improvement Score of +1 at Any Reviewed Visit by Age: C-DM and C-HT Cohorts mITT Population

	Age <= 48 years (N=31)	Age > 48 years (N=15)	Overall (N=46)
Subjects who had a Median Clinical Improvement of +1 at Any Reviewed Visit			
C-DM Cohort			
Responder n (%)	16 (51.6%)	7 (46.7%)	23 (50.0%)
Non-Responder n (%)	2 (6.5%)		2 (4.3%)
C-HT Cohort			
Responder n (%)	10 (32.3%)	7 (46.7%)	17 (37.0%)
Non-Responder n (%)	3 (9.7%)	1 (6.7%)	4 (8.7%)
Overall			
Responder n (%)	26 (83.9%)	14 (93.3%)	40 (87.0%)
Non-Responder n (%)	5 (16.1%)	1 (6.7%)	6 (13.0%)

Note: 1. C-DM group includes subjects with diabetes mellitus type 2 and/or impaired glucose tolerance at screening and Day 1 as determined by two or more abnormal oGTTs. C-HT group includes subjects with diagnosis of hypertension at screening but without diabetes mellitus type 2 and/or impaired glucose tolerance.
2. Age Group is broken by tertiles. Low and medium = Age Group 1 (Age <= 48 years), High = Age Group 2 (Age > 48 years).
Source: Listings 16.2.6.3, 16.2.4.1

Number of Subjects With or Without a Median Clinical Improvement Score of +1 at Any Reviewed Visit by Sex: C-DM and C-HT Cohorts
mITT Population

	Male (N=14)	Female (N=32)	Overall (N=46)
Subjects who had a Median Clinical Improvement of +1 at Any Reviewed Visit			
C-DM Cohort			
Responder n (%)	5 (35.7%)	18 (56.3%)	23 (50.0%)
Non-Responder n (%)	1 (7.1%)	1 (3.1%)	2 (4.3%)
C-HT Cohort			
Responder n (%)	6 (42.9%)	11 (34.4%)	17 (37.0%)
Non-Responder n (%)	2 (14.3%)	2 (6.3%)	4 (8.7%)
Overall			
Responder n (%)	11 (78.6%)	29 (90.6%)	40 (87.0%)
Non-Responder n (%)	3 (21.4%)	3 (9.4%)	6 (13.0%)

Note: 1. C-DM group includes subjects with diabetes mellitus type 2 and/or impaired glucose tolerance at screening and Day 1 as determined by two or more abnormal oGTTs. C-HT group includes subjects with diagnosis of hypertension at screening but without diabetes mellitus type 2 and/or impaired glucose tolerance.
Source: Listings 16.2.6.3, 16.2.4.1

4.2 Other Special/Subgroup Populations

None of much importance was done.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Although the sponsor could achieve his goal of showing statistical evidence that the test drug provided at least 20% more response than non-response (20% margin) with respect to AUC-Glucose in the C-DM cohort, clinical judgment should be given priority in this open-label study with titration and meager data. More details on statistical issues are in the “CHECK LIST” at the very end.

There was no evidence of diastolic blood pressure lowering in the C-HT cohort.

Template points:

- breaking the blind - It was an open-label study.
- unblinded or unplanned interim analyses - None
- high percentage of dropouts – Yes, 32%. But not unusual
- inappropriate imputation for missing values – deferred to clinical (details of handling missing values are in the “CHECK LIST” at the very end.

- change of primary endpoint during conduct of the trial – other changes but not primary endpoint
- dropping/adding treatment arms – N/A
- sample size modification – No.

> "The sample size of 50 subjects was chosen by clinical judgment. Published case reports suggest that mifepristone can reverse many of the signs and symptoms of hypercortisolemia. However, there have been no prospective clinical trials from which to estimate a treatment effect or standard deviation of treatment effect in subjects with Cushing's Syndrome.
>
> Prior to Amendment 5, the protocol stated that sample size would be re-estimated by the Data Review Board by estimating conditional power after 15 C-DM and 10 C-HT subjects had completed the study. However, the pattern of subject enrollments into the study is such that the study will be fully enrolled (50 subjects) prior to reaching the predefined point of sample size re-estimation. Therefore, the Data Review Board will not re-estimate the sample size."

- inconsistency of results across subgroups – Numerically, yes but statistical conclusions are not safe because of meager data.
- Type I error inflation due to multiplicity – Yes, due to two cohorts - C-DM and C-HT.

Comment: If the two cohorts were studied with different study names, we would not think about multiplicity.

- Planned and unplanned adaptations - No
- Non-Inferiority – No (no commercial comparator was available).

5.2 Conclusions and Recommendations

Although the sponsor could achieve his goal of showing statistical evidence that the test drug provided at least 20% more response than non-response (20% margin) with respect to AUC-Glucose in the C-DM cohort, clinical judgment should be given priority in this open-label study with titration and meager data. More details on statistical issues are in the "CHECK LIST" at the very end.

There was no evidence of diastolic blood pressure lowering in the C-HT cohort.

Labeling

Though no statistical significance was claimed for the secondary efficacy variables, there is one danger that non-statisticians may not be fully alert that these descriptive statistics do not mean much. They are just numerical results based on one sample; there is no assurance or confidence regarding the population or the reality.

The definition of a Responder in the key secondary efficacy variable: "A responder was defined as a subject whose median reviewer score was + 1 at any reviewed visit after baseline through Week 24/ET" with the phrase "at any reviewed visit," gives multiple opportunities for a success

and is not as dependable as a response at any one time-point. Therefore, the results of this key variable, if are allowed to be in the labeling at all, this point should be emphasized.

APPENDICES (Add When Needed)

Appendix I, LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

AAG	α1-acid-glycoprotein
ACTH	Adrenocorticotropic hormone
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AUC	Area under the concentration-time curve
$AUC_{glucose}$	Area under the concentration-time curve for glucose
$AUC_{insulin}$	Area under the concentration-time curve for insulin
BCE	Bone collagen equivalents
BDI-II	Beck Depression Inventory II
BMD	Bone mineral density
BMI	Body mass index
bpm	Beats per minute
BSAP	Bone-specific alkaline phosphatase
BUN	Blood urea nitrogen
CBC	Complete blood count
C-DM	Subjects with Cushing's syndrome and diabetes or impaired glucose tolerance
C-HT	Subjects with Cushing's syndrome and hypertension
CI	Confidence interval
CRA	Clinical Research Associate
CREST	calcinosis, Raynaud phenomenon, eosphageal dysmotility, sclerodactyly, and telangiectasia
CRF	Case report form
CRH	Corticotropin-releasing hormone
CRO	Contract research organization
CYP	Cytochrome P450
d	Day

D&C	Dilation and curettage
DPP-IV	dipeptidyl peptidase-4
DRB	Data Review Board
DXA	Dual-energy x-ray absorptiometry
ECG	Electrocardiogram
ET	Early termination
F	Female
GABA	Gamma aminobutyric acid
GAGS	Global Acne Grading System
GCP	Good Clinical Practice
HDL	High-density lipoprotein
HbA1c	Hemoglobin A1c (glycosylated hemoglobin)
HIV	Human immunodeficiency virus
HOMA-IR	Homeostatic model of insulin resistance
HPA	Hypothalamic-pituitary-adrenal
HR	Heart rate
HTN	Hypertension
ICF	Informed consent form
ICH	International Conference on Harmonization
IEC	Independent Ethics Committee
IPSS	Inferior petrosal sinus sampling
IRB	Institutional Review Board
ITT	Intent-to-Treat (population)
IUD	Intrauterine device
LDL	Low-density lipoprotein
LLN	Lower limit of normal
LSC_{95}	Least significant change 95; amount of bone needed to be 95% certain an individual change is real
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities

Min	Minimum
min	Minute
mITT	Modified Intent-to-Treat (population)
MRI	Magnetic resonance imaging
NSAID	Non-steroidal anti-inflammatory drug
NTx	N-telopeptide of type 1 collagen
oGTT	Oral glucose tolerance test
PAEC	Progesterone modulator-associated endometrial changes
PF	Physical Function
PK	Pharmacokinetic(s)
PMD	Psychotic major depression
PPARγ	Proliferator-activated receptor gamma
PRN	As needed
PT	Preferred term
QD	Once daily
QTc	Corrected QT interval
QTcB	Corrected QT interval (Bonferroni's method)
QTcF	Corrected QT interval (Fridericia's method)
QOL	Quality of Life
RBC	Red blood cell
RRSEC	RR interval in seconds (electrocardiogram measurement)
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Standard deviation
SE	Standard error (of the mean)
SERM	Selective estrogen receptor modulator
SF-36	Short Form-36 Health Survey
sqrt	Square root
SSRI	Selective serotonin reuptake inhibitor

T4	Thyroxine
TAT	Thrombin-antithrombin
TEAE	Treatment-emergent adverse event
TSH	Thyroid-stimulating hormone
TVUS	Transvaginal ultrasound
UFC	Urinary free cortisol
ULN	Upper limit of normal
US	United States
VLDL	Very-low-density lipoprotein
w	Week
WBC	White blood cell
WHO	World Health Organization

Japobrata Choudhury, Ph.D.
Mathematical Statistician

Concur: Dr. Sahlroot

CC:
Archival NDA 202107

HFD-510/Dr. Parks
HFD-510/Dr. Roman
HFD-510/Dr. Zemskova
HFD-700/ Ms. Patrician
HFD-715/Dr. Permutt
HFD-715/Dr. Sahlroot
HFD-715/Dr. Choudhury

J.Choudhury:6-1184: 1/19/12

This review consists of 46 pages.

CHECK LIST

Number of Pivotal Studies: One

Trial Specification
Specify for each trial:

Protocol Number (s): CI073-400
Protocol Title (optional): Open-label Study of the Efficacy and Safety of CORLUX® (mifepristone) in the Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome
Phase: 3
Control: Own baseline
Blinding: Open-Label
Number of Centers: 20 investigators enrolled patients
Region(s) (Country): US
Duration: 24 Weeks
Treatment Arms: Mifepristone (test drug)
Treatment Schedule: 300 mg mifepristone tablets administered orally once daily, titrated as needed

Randomization: No
Ratio: N/A
Method of Randomization: N/A

Primary Endpoints:

(1) The primary endpoint for subjects with Cushing's syndrome and diabetes mellitus (or impaired glucose tolerance) (C-DM cohort) was the change in the area under the concentration-time cure for glucose (AUCglucose) in the 2-hour oral glucose tolerance test (oGTT) from baseline to Week 24 in subjects with diabetes/impaired glucose tolerance with or without hypertension at screening.

(2) Changes in diastolic blood pressure were analyzed for subjects with Cushing's syndrome and a diagnosis of hypertension (C-HT) at screening (ie, without impaired glucose tolerance or diabetes) as a primary endpoint.

Primary Analysis Population: (e.g., ITT, mITT, Per-Protocol…)

Modified-Intent-to-Treat Population

The efficacy population is defined as all subjects who received a total of at least 30 days of mifepristone during the 24-week treatment period. The primary and secondary endpoints will be evaluated using this modified Intent-to-treat population (mITT). The 30 days of treatment do not need to be consecutive. Subjects who receive >=30 days of mifepristone but terminate prior to

week 24 are included in the mITT population. The primary endpoint and secondary efficacy analyses will exclude the following subjects:

C-DM:
Did not undergo at least one oGTT at baseline (day 1) or did not have any oGTT with a valid $AUC_{glucose}$ measurement after day 1 (see Section 5.6 for the definition of valid $AUC_{glucose}$ measurement).

C-HT:
Did not have blood pressure measured (or measured incorrectly) at baseline (day 1) or did not have any valid measurement after day 1.

Statistical Design: Superiority/ Non-Inferiority
Comparison with own baseline
Adaptive Design: No

Primary Statistical Methodology:

(1) A responder analysis using the modified Intent-to- Treat (mITT) population was used to measure success on this primary efficacy endpoint. A responder was defined as a subject who experienced at least a 25% decrease in AUCglucose from baseline to Week 24. This efficacy measurement was to be declared successful if the lower limit of the exact 95% binomial confidence interval for the responder rate was >=20%.

(2) A responder analysis of the reduction in diastolic blood pressure from baseline to Week 24 was performed for the efficacy population. A responder was defined as a subject who experienced a >= 5 mmHg decline in diastolic blood pressure from baseline to Week 24. This efficacy measurement was to be declared successful if the lower limit of the exact 95% binomial confidence interval for the responder rate was >= 20%.

The study was considered to have had a positive outcome and to have achieved the primary endpoint if either of the two primary measures described above were positive.

Interim Analysis: No

Sample Size:
The sample size of 50 subjects was chosen by clinical judgment. Published case reports suggest that mifepristone can reverse many of the signs and symptoms of hypercortisolemia. However, there have been no prospective clinical trials from which to estimate a treatment effect or standard deviation of treatment effect in subjects with Cushing's Syndrome.

Prior to Amendment 5, the protocol stated that sample size would be re-estimated by the Data Review Board by estimating conditional power after 15 C-DM and 10 C-HT subjects had completed the study. However, the pattern of subject enrollments into the study is such that the study will be fully enrolled (50 subjects) prior to reaching the predefined point of sample size re-estimation. Therefore, the Data Review Board will not re-estimate the sample size.

- Was there an **Alternative Analysis** in case of violation of assumption; e.g., Lack of normality, Proportional Hazards Assumption violation. N/A (responder analysis)

- Were there any major changes, such as changing the statistical analysis methodology or changing the primary endpoint variable? No [Yes, as mentioned before "Results and Conclusions" but none was major]

- Were the **Covariates** pre-specified in the protocol? No covariates in the primary analyses
- Did the Applicant perform **Sensitivity Analyses**? No

- How were the **Missing Data** handled?

> "**$AUC_{glucose}$**
>
> $AUC_{glucose}$ measures will be obtained for all subjects (C-DM and C-HT) at Baseline and Weeks 6, 10, 16, 24 (or early termination visit).
>
> When calculating the AUC at a given time point, the following rules for handling missing data will be applied:
>
> . If data for fasting plasma glucose (time point 0) or the plasma glucose 30 minute post oral glucose administration time point are missing, no AUC calculation will be performed for that particular visit and the AUC for that visit will be counted as missing.
>
> . If glucose concentration data are not available for more than one oGTT (post oral glucose administration) time point (i.e., 30, 60, 90 and 120 minutes) for a particular visit, then the $AUC_{glucose}$ will not be calculated for that visit.
>
> . If data for only one plasma glucose (post oral glucose administration) time point other than the 30 minute time point (i.e., either 60 or 90 minutes) is missing, then AUC will be calculated using available data. For example, if 90 minute value is missing, then a larger trapezoid will be constructed to connect the 60 min time point to the 120 min time point.
>
> . If only the 120 minute plasma glucose is missing, the available time points will be used to calculate $AUC_{glucose}$. In such a case, the 120 minute plasma glucose for the baseline oGTT will be disregarded. $AUC_{glucose}$ will be calculated using time points 0, 30, 60, 90 minutes; $AUC_{glucose\ 0\text{-}90}$ will be used for both baseline and final observation.
>
> When evaluating the primary endpoint among C-DM subjects, an endpoint analysis will be calculated using the change from baseline $AUC_{glucose}$ to the last value for $AUC_{glucose}$ obtained (week 24 or early termination visit), provided that that last observation occurred no more than 14 days after the patient stopped taking the medication. The 14 day period of time has been chosen because, based on the half-life of mifepristone, 14 days exceeds the expected duration of clearance of the drug from the circulation in the vast majority of people who take the drug. Thus the 14 day limit has been chosen to best reflect the effect of the drug on $AUC_{glucose}$ in situations where the interval between discontinuation of study drug and the last oGTT is prolonged. In the case where the last

observation occurred more than 14 days after the patient stopped taking the medication, the most recent prior value of $AUC_{glucose}$ will be used.

Blood Pressure
Blood pressure evaluations will be taken for all subjects (C-DM and C-HT) at screening, baseline, days 7, 14, and 28, and weeks 6,8,10,12,16, 20, and 24 (or early termination).

For the primary endpoint evaluation ofC-HT subjects, an endpoint analysis will be calculated using the change from baseline diastolic blood pressure to the last valid value of diastolic blood pressure obtained (week 24 or early termination visit), provided the last observation occurred no more than 14 days after the patient stopped taking the medication. In the case where the last observation occurred more than 14 days after the patient stopped taking the medication, the most recent prior value of diastolic blood pressure will be used."

- Was there a **Multiplicity** involved?
  If yes,
  Multiple Arms (Yes/No)? No
  Multiple Endpoints (Yes/No)? No
  - Which method was used to control for type I error? None
    - Type I error inflation due to multiplicity – Yes, due to two cohorts - C-DM and C-HT.

Comment: If the two cohorts were studied with different study names, we would not think about multiplicity.

- **Multiple Secondary Endpoints**: Are they being included in the label? If yes, method to control for type 1 error.

No statistical significance was claimed for the secondary efficacy variables. There is one danger that non-statisticians may not be fully alert that these descriptive statistics do not mean much. They are just numeric results based on one sample; there is no assurance or confidence regarding the population or the reality.

Were Subgroup Analyses Performed (Yes/No)? Yes, some, including Gender and Age.

- Were there any **Discrepancies** between the protocol/statistical analysis plan vs. the study report?

As mentioned before "Results and Conclusions," there were some changes made but none were major.

- Overall, was the study positive (Yes/No)? Statistically, yes for C-DM cohort.

Although the sponsor could achieve his goal of showing statistical evidence that the test drug provided at least 20% more response than non-response (20% margin) with respect to AUC-Glucose in the C-DM cohort, clinical judgment should be given priority in this open-label study with titration and meager data.

There was no evidence of diastolic blood pressure lowering in the C-HT cohort.

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

/s/

JAPOBRATA CHOUDHURY
01/19/2012

JON T SAHLROOT
01/19/2012
concur

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

STATISTICAL REVIEW AND EVALUATION

ADDENDUM

NDA:	**202107**
Drug Name:	**CORLUX® (mifepristone) 300 tablets**
Indication:	**Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome**
Applicant:	**Corcept Therapeutics** **149 Commonwealth Drive** **Menlo Park, California 94025**
Date(s):	**Study Data Submitted: 17 May 2011** **To Reviewer: 13 December 2011**
Review Priority:	**Standard**
Biometrics Division:	**Division 6**
Statistical Reviewer:	**Steve Thomson**
Concurring Reviewers:	**Karl Lin, Ph.D.**
Medical Division:	**Metabolism and Endocrinology Products**
Toxicologist Team:	**Patricia Brundage, Ph.D.** **Todd Bourcier, Ph.D.**
Project Manager:	**Jena Weber, BS**

Keywords: Carcinogenicity

When the original submission was reviewed, the FDA Executive Carcinogenicity Assessment Committee (ECAC) requested further analysis of the tumors in the mammary glands of female rats, particularly tests of trend over the pooled controls, low, and medium dose groups. The statistical analyses of carcinogenicity use the Haseman-Lin multiplicity adjustments, i.e., for a rough 10% false rejection error rate the usual significance levels for the test of overall trend are compared to 0.005 for common tumors and 0.025 for rare tumors. Pairwise comparisons between the high dose and controls are compared to 0.01 for common tumors and 0.05 for rare tumors. As noted elsewhere, using this rule for the medium or low dose group can be expected to increase the false rejection error rate to some value larger than the rough 10%.

In particular, the ECAC expressed some question about the statistical significance levels of the tests of adenoma, and pooled adenoma/adenocarcinoma comparing the medium dose group to pooled controls ($p = 0.0022$, $p < 0.0001$, both ≤ 0.01). Note that the high statistical significance of these tests is due to the relatively high tumor incidence in the medium dose group compared to the pooled controls. Treating the middle dose group as the highest dose for tests of trend and pairwise differences (and thus inflating the multiplicity adjusted significance level to some value above the nominal 10% level) , the test of trend in adenoma over the pooled controls, low, and medium dose groups (denoted "C-M trend" below) is quite close to significance ($p = 0.0057 \approx 0.005$) while the test of trend in pooled adenoma/ adenocarcinoma would be significant ($p < 0.0001 < 0.005$). However, the similar test of trend in adenocarcinoma would not be significant ($p = 0.0192 > 0.005$).

Table Addendum.1: Incidence and Significance Levels of Selected Tests for Neoplasms in Female Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Significance Levels C-H Trend	C-M Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
MAMMARY GLAND										
# Evaluated	59	60	59	59	60					
Adenocarcinoma	17	16	17	25	17	.7056	.0192	.6618	.0285	.4354
Adenoma	9	12	23	24	16	.5649	.0057	.1942	.0022	.0020
Adenoma/Adenocarcinoma	22	22	32	42	26	.6871	<.0001	.3527	<.0001	.0140
Fibroadenoma	22	17	10	6	1	1	.9998	1	.9999	.9938

The basis of the poly-k test is to down-weight those animals that die early without a tumor whose incidence is being analyzed. The k usually used for such tests is k=3, which it is claimed fits the typical profile of tumor incidence over time. For example, with k=3, an animal that dies halfway through a study without the particular tumor being assessed counts as only 0.125 of an animal in the computations for that particular tumor (i.e. $(1/2)^k = 1/8 = 0.125$ for k=3). From, Table 2, below it is clear that there is no dramatic reduction in the adjusted number of animals at risk over dose. So the inconsistency in results for the medium dose group and the high dose group does not seem to be due early deaths without tumor in the high dose group.

Table Addendum.2: Mortality Adjusted Number of Animals at Risk

Organ/Tumor	C1+C2	Low	Med	High
MAMMARY GLAND				
Adenocarcinoma	98	47	49	54
Adenoma	96	49	53	54
Adenoma/Adenocarcinoma	101	50	54	54

There do seem to be weight gain differences between the high dose group and the medium dose group. Thus, one might speculate that the differences in tumor incidence may be one of those artifactual results that occur in real studies or it may be that the lower weight gain in the high dose group is sufficient to counter any carcinogenic effect associated with the highest dose. However, such an evaluation requires the expertise of the toxicologist,

---

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

/s/

STEVEN F THOMSON
01/18/2012

KARL K LIN
01/18/2012
Concur with review

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA:	**202107**
Drug Name:	**CORLUX® (mifepristone) 300 tablets**
Indication:	**Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome**
Applicant:	**Corcept Therapeutics** **149 Commonwealth Drive** **Menlo Park, California 94025**
Date(s):	**Study Data Submitted: 17 May 2011** **To Reviewer: 2 June 2011**
Review Priority:	**Standard**
Biometrics Division:	**Division 6**
Statistical Reviewer:	**Steve Thomson**
Concurring Reviewers:	**Karl Lin, Ph.D.**
Medical Division:	**Metabolism and Endocrinology Products**
Toxicologist Team:	**Patricia Brundage, Ph.D.** **Todd Bourcier, Ph.D.**
Project Manager:	**Jena Weber, BS**

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

Table of Contents

1. EXECUTIVE SUMMARY

Reports from two studies, in rats and mice, were provided. The rat study was conducted by (b) (4) This report states that the "objective of this study was to investigate the carcinogenic potential of [compound] C-1073 [, i.e. Corlux,] following daily oral administration by gavage for 104 consecutive weeks." (page 17 of report) The mouse study was conducted (b) (4), and its report notes that the "study was conducted . . . to evaluate the oncogenic potential of [Corlux] following oral administration to mice for 104 weeks." (page 9 of report)

1.1. Conclusions and Recommendations

In both studies treatment was applied by oral gavage with 0.25% carboxymethylcellulose and 0.2% Tween® 80 dissolved in sterile water as the vehicle. Dosing with the vehicle alone or test article was administered to all groups once daily at a dose volume of 10mL/kg/dose. Other gross aspects of the study designs for the main study animals are summarized below:

The rat study was conducted at the (b) (4)

Table 1. Design of Albino Rat Study

Treatment Group	# Animals	Dosage (mg/kg/day)
1. Vehicle	60	0
2. Vehicle	60	0
3. Low	60	5
4. Medium	60	25
5. High	60	125

The mouse study was conducted at the (b) (4)

Table 2. Design of Mouse Study

Treatment Group	# Animals	Dosage (mg/kg/day)	
		Males	Females
1. Vehicle	60	0	0
2. Low	60	12.5	25
3. Medium	60	65	100
4. High	60	125	300/200/125[1]

[1]Females received 300 mg/kg from Weeks 1 to 35, 200 mg/kg from Weeks 36 to 53, and 125 mg/kg from Weeks 54 to 104.

More detailed descriptions of the studies are provided in Section 3.2.1 and 3.2.2 below. In this report the vehicle groups are sometimes referred to as "control groups" while the other dose groups are referred to as "actual dose groups" or "treated groups." Simple summary life tables in mortality are presented in the report in these sections of the report.

In Appendix 1, Figures A.1.1 and A.1.2 for rats and Figures A.1.3 and A.1.4 for mice in display Kaplan-Meier estimated survival curves for each study group for each species and gender combination. The results of tests of trend and differences in survival are displayed in Tables 3 and 4 below:

Table 3. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Rat Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1&2, 3-5	0.0206	0.0182	0.0815	0.0708
Homogeneity over Groups 1-5	0.0375	0.0345	0.0889	0.0838
No trend over Groups 1&2, 3-5	0.0081	0.0087	0.0129	0.0076
No Difference Between Groups 1&2 vs 5	0.0020	0.0019	0.0086	0.0067

From Figure A.1.1, in male rats, at first there is a clear roughly decreasing survival over dose with the high dose group having the highest overall mortality. This is consistent with the results of the test above. For example, the tests of trend in mortality over the pooled vehicle groups with dosing groups 3-5 is highly statistically significant (logrank p = 0.0081, Wilcoxon p = 0.0087), confirmed by the pairwise test between the high dose and pooled controls (logrank p = 0.0020, Wilcoxon p = 0.0019). The tests of no homogeneity over all three dosing groups and pooled controls were statistically significant (logrank p = 0.0206, Wilcoxon p = 0.0182), as was the less appropriate test over all five groups, distinguishing between vehicle controls (logrank p = 0.0375, Wilcoxon p = 0.0345). In female rats results are quite inconsistent with these results. The high dose group has the highest survival, with vehicle group 2 having the lowest, and the other groups somewhat intertwined. This is sufficient to provide statistically significant evidence of a trend (logrank p = 0.0129, Wilcoxon p = 0.0076), but, again, the trend is for increasing survival over increasing dose. The difference between the high dose and pooled vehicle was also statistically significant (logrank p = 0.0086, Wilcoxon p = 0.0067). In female rats, the tests of no homogeneity over all five study groups were not quite statistically significant (logrank p = 0.0889, Wilcoxon p = 0.0838), as were the more appropriate tests of no homogeneity using the pooled controls (logrank p = 0.0815, Wilcoxon p = 0.0708).

Table 4. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Mouse Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Mouse Homogeneity over Groups 1-4	0.1741	0.0482	0.0005	<0.0001
No trend over Groups 1-4	0.0677	0.0240	0.0002	<0.0001
No Difference Between Groups 1 vs 4	0.0364	0.0090	0.0021	0.0001

From Figure A.1.3, in male mice, the high dose group has the highest overall mortality. The low dose group more or less has the next highest mortality, particularly in the middle of the study, with the vehicle and mid dose group mostly intertwined. This is consistent with the statistically significant Wilcoxon statistic (p = 0.0240), while the logrank test is not quite

significant at the commonly accepted 0.5 level (p = 0.0677). As noted in section 1.2.2 this is because the log rank test is more sensitive to later differences than early differences in survival. Note there is statistically significant evidence of a difference between the high dose and vehicle (logrank p = 0.0364, Wilcoxon p = 0.0090). The somewhat inconsistent results in the outcomes of tests of survival differences continue when testing homogeneity in survival over all four groups in male mice (logrank p = 0.1741, Wilcoxon p = 0.0482). Results are somewhat different in female mice, especially in terms of statistical significance. Again, from Figure A.1.4, in female mice, the high dose group has the highest overall mortality, quite separated from the results in other study groups. Relatively speaking, survival in the other study groups tends to be somewhat intertwined, although again the low dose group tends to have somewhat higher mortality in the middle of the study period. The much higher mortality in the high dose group is sufficient to explain the highly statistically significant tests for overall homogeneity, trend, and pairwise difference between the high dose and controls (all 3 logrank $p \leq 0.0021$, all 3 Wilcoxon $p \leq 0.0001$).

The results above do differ from those of an experimental Bayesian nonparametric analysis of survival as described in Appendix 2. The estimated survival curves do seem to be consistent with the conclusions above. However, the analysis of parameterized differences seems to suggest that there is no overwhelmingly strong evidence of differences in either gender or species between the possibly pooled vehicle and the other treatment groups.

Of course in a carcinogenicity study, primary interest is on the occurrence of cancers. Statistical analysis compares tumor incidence over dose groups. Complete tumor incidence tables for each organ are provided in Tables A.3.3 through A.3.6 in Appendix 3. Tables 5 and 6 below display those organ tumor combinations that had at least one test of trend or pairwise difference from control that was statistically significant at the usual 0.05 level. For each species by gender by organ the number of animals analyzed and used in the statistical tests is presented first. The tumor incidence for each organ is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high, medium, and low dose groups. These statistical tests are conditioned on the animals actually evaluated, ignoring those not analyzed. For tests in rats the two control groups are pooled.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.5, below, are often applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors (incidence > 1%) and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. As also discussed in section 1.3.1.4, using these adjustments for other tests, like the trend over the vehicle, low, and medium dose groups in mice and the pairwise comparisons between the vehicle and the medium-high, medium, and low dose groups can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate. The period '.' in these tables denotes the p-values of tests of dose groups with no tumors in any group.

Table 5. Potentially Statistically Significant Neoplasms in Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Significance Levels Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
Male									
HEMOLYM. TISSUE									
# Evaluated	60	60	60	60	60				
Malignant lymphoma	0	0	0	1	2	.0265	.0914	.3137	.
MAMMARY GLAND									
# Evaluated	50	48	54	56	54				
Adenoma/Adenocarcinoma	0	0	3	0	0	.7875	.	.	.0388
PITUITARY									
# Evaluated	60	60	60	60	60				
Adenoma: pars distalis	29	29	31	37	33	.1314	.1088	.0408	.2215
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Foll. cell	1	1	0	2	3	.0462	.1595	.3722	1
Adenoma: follicular cell	1	0	0	1	3	.0213	.0809	.5304	1
Female									
KIDNEY									
# Evaluated	60	60	60	60	60				
Adenoma: tubular cell	0	0	0	0	2	.0480	.1320	.	.
LIVER									
# Evaluated	60	60	60	60	60				
Adenoma: hepatocellular	0	1	1	3	6	.0038	.0098	.1165	.5635
MAMMARY GLAND									
# Evaluated	59	60	59	59	60				
Adenocarcinoma	17	16	17	25	17	.7056	.6618	.0285	.4354
Adenoma	9	12	23	24	16	.5649	.1942	.0022	.0020
Adenoma/Adenocarcinoma	22	22	32	42	26	.6871	.3527	<0.0001	.0140
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Foll. cell	0	1	3	1	11	<0.0001	.0001	.5698	.1122
Adenoma: follicular cell	0	1	2	1	8	.0004	.0014	.5698	.2641
Carcinoma: follicular cell	0	0	1	0	3	.0187	.0471	.	.3381

Using the incidence in the pooled control group to specify whether a tumor is treated as common or rare, in both male and female rats, the test of trend in follicular cell adenoma of the thyroid is statistically significant (Males: $p = 0.0213$, Females: $p = 0.0004$, both < 0.025). In female rats the tests of trend in follicular cell carcinoma and pooled adenoma/carcinoma of the thyroid were both also statistically significant ($p = 0.0187$, $p < 0.0001$, respectively, both < 0.025). Further, the pairwise comparisons between the pooled controls and high dose group in these three neoplasms in females were also statistically significant ($p = 0.0014$, 0.0471, $0.0001 < 0.05$). In female rats the test of trend in hepatocellular adenoma in the liver was also statistically significant ($p = 0.0038 < 0.025$) as was the pairwise comparison between the high dose group and pooled controls ($p = 0.0098 < 0.05$). Although this may extend the multiplicity correction into a region where one arguably should not go, the pairwise comparison between the medium dose groups versus the pooled controls in terms of pooled adenoma/adenocarcinoma of the mammary gland in female rats was also statistically significant ($p < 0.0001 < 0.01$). With the same caveat, the pairwise comparisons between controls and the medium and low dose groups in terms of mammary adenoma could also be classified as statistically significant ($p = 0.0022$ and $0.0020 < 0.01$, respectively). In male rats the test of trend in malignant lymphoma in

hemolymphatic tissue was close to the multiplicity adjusted level of statistical significance ($p = 0.0265 \approx 0.025$). Finally, the count of three male rats in the low dose group with pooled adenoma and adenocarcinoma of the mammaries was sufficient to make the p-value of the test with the pooled controls small enough to be nominally significant in this extended criteria ($p = 0.0388 < 0.05$). No other comparisons reached the multiplicity adjusted statistical significance levels.

Table 6. Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

	Incidence				Significance Levels			
Organ/ Tumor	Cntr1	Low	Med	Hi	Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
liver								
# Evaluated	60	60	60	60				
adenoma, hepatocellular[B]	1	0	5	2	.0334	.2637	.0914	1
lung								
# Evaluated	60	60	60	60				
Adenoma/Carcinoma Bronch. Alv.	10	7	9	10	.0370	.0857	.6783	.8188
carcinoma, bronch. alv.[M]	2	1	0	4	.0336	.0932	1	.8634

If we use the incidence in the control group no treatment group to specify whether a tumor is treated as common or rare, in both male and female mice no comparisons achieved the levels of multiplicity adjusted significance. In males no tests even achieved the nominal 0.05 level.

Complete incidence tables in both species are provided in tables A.3.3 through A.3.6 in Appendix 3.

An alternative Bayesian analyses of tumorigenicity is presented in Appendix 4. It suggests that in female rats the probability of a dose related increasing likelihood of hepatocellular adenoma of the liver, follicular cell adenoma of the thyroid and pooled follicular cell adenoma and carcinoma of the thyroid are all about 0.99 or more. However, none of the other tumors analyzed in either gender in either species were associated with very strong evidence of an increasing dose effect.

1.2. Brief Overview of the Studies

This submission had an oral study in albino rats:

(b) (4) **Study # 77389 Title: A 2-Year Oral Gavage Carcinogenicity Study of C-1073 in the Albino Rat,**

and a similar, mouse study:

(b) (4) **Study # 950-005 Title: 104-Week Oral Oncogenicity Study of C1073 in Mice.**

Somewhat detailed descriptions of these studies are available in Sections 3.2.1 and 3.2.2, below.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

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

1.3.1.1. Dual Controls:

In the rat study, the Sponsor provides two supposedly identical vehicle control groups. All tables and plots in this report distinguish between the two control groups, groups 1 and 2. Prior to tests the Sponsor tests for differences in the control groups and provides different tests on the basis of these tests. The first issue with such a procedure is that results of tests on treatment groups are conditional on the outcomes of the tests between the controls, whereas the significance values are computed assuming the tests are not conditional. Thus the distributional assumptions of the usual unconditional tests are not met. Also, of more importance is that unless there are systemic problems with the conduct of the study, any observed differences should be due to random fluctuations between the treatment groups. That is, pre-study randomization to two identical controls should be equivalent to post-study randomization into two control groups. In the latter circumstances it would seem that few analysts would place any weight on observed differences between the control groups (since a simple rerandomization would almost surely eliminate any differences). But then logically no weight should be placed on any observed differences between vehicle controls in the current studies, and on differing results when control groups are tested against other treatment groups. Finally, note that this procedure increases the number of statistical tests, and thus increases the probability of a false conclusion of treatment differences. Hence, this reviewer would argue against the separate analyses as conducted by the Sponsor, and all tests in the FDA analysis, both tests of differences in survival and tests of differences tumorigenicity use a single pooled control group and ignore possible differences in controls.

1.3.1.2. Survival Analysis:

The survival analyses presented here are based on both the log rank test and the Wilcoxon test comparing survival curves. The log rank tests tend to put higher weight on later events, while the Wilcoxon test tends to weight events more equally, and thus is more sensitive to earlier differences in survival. The logrank test is most powerful when the survival curves track each other, and thus the hazards, i.e., the conditional probability of the event in the next infinitesimal interval, would be roughly proportional. This is the test used by the Sponsor. In the FDA analysis, both tests were used to test both homogeneity of survival among the treatment groups and the effect of dose on trend in survival. Appendix 1 reviews the specific animal survival analyses in more detail. The results of the Sponsor's analysis are summarized in Sections 3.2.1.1 and 3.2.2.1. An experimental Bayesian nonparametric analysis of survival is given in Appendix 2. However, these Bayesian results are inconsistent with the tests above. This is an issue that requires further study.

1.3.1.3. Multiplicity of Tests on Survival:

Using the logrank and Wilcoxon tests, for each gender there are eight tests of survival in rats and six tests of survival in mice for each species by gender combination. If we were to assume that any set of tests are independent across comparisons, which clearly they are not, and assume that there is absolutely no difference in survival, the probability of at least one statistically significant result in each gender, at the usual 0.05 level, is about 0.265, and about 0.460 of at least one statistically significant result in at least one gender. Such is the possible price paid for the multiplicity of hypothesis tests in the frequentist paradigm.

1.3.1.4. Tests on Neoplasms:

The Sponsor's analyses use Cochran-Armitage tests of trend and Peto analyses of neoplasms, The analyses in this report are based on poly-k analysis of tumor incidence. The poly-k test is a modification of the original Cochran-Armitage test of trend in response to dose, adjusted for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that the poly-k modification of the Cochran-Armitage tests of trend has been recommended over the corresponding Peto tests.

Appendix 4 provides an alternative Bayesian assessments of the effect of trend in dose on tumor incidence. The Bayesian approach takes the original probability model for the observations approach and uses probability as a measure of ignorance about the exact locations of parameters in the probability model. Then the observed data are used to reduce that ignorance about the parameters.

1.3.1.5. Multiplicity of Tests on Neoplasms:

Frequentist hypothesis testing involves accepting or rejecting hypotheses about the parameters of interest on the basis of the values of some statistic. If one does not provide some sort of multiplicity adjustment to the significance level, the chances of rejecting one or more true null hypothesis increases as the number of such tests increases. To avoid this it is common to adjust for multiplicity in hypothesis testing resulting in an adjustment in experiment-wise Type I error (i.e., the probability of rejecting a true null hypothesis). Based on his extensive experience with such carcinogenicity analyses in standard laboratory rodents, for pairwise tests between the high dose group and controls in two species, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Similarly, Lin and Rahman (1998) showed that tests of trend should be tested at a 0.025 level for rare tumors and 0.005 for common tumors. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

Significance levels of the pairwise tests between the vehicle groups and the low and medium dose groups are also provided. Including these tests can be expected to increase the overall type I error rate to some level above the rough 10% level. Even if one uses the Haseman-Lin-Rahman rules above, the overall type I error associated with including the tests

between the vehicle and the low and medium-low dose groups may be considerably larger than the rough 10% when these rules are restricted to the test of trend and pairwise differences between the high dose and vehicle.

The Bayesian approach is based on hierarchical models, where parameters for a specific tumor are adjusted for the other tumors, and thus adjust for multiplicity.

1.3.1.6. Validity of the Designs:

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

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

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

Lin and Ali (2006), quoting work by Haseman, have suggested that in standard laboratory rodent species, a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. Note that as a percentage of animals that survived to week 91, this criterion is not met in the high dose group in either gender. (Please see tables 8 and 9 on page 14). Like the other comments in this section this requires the expertise of the toxicologist, this does suggest that the MTD was exceeded.

The mean weight values and derived differences and ratios in the following table were taken directly from the Sponsor's reports (Rat Tables 1, pages 40-47, Mouse Table 2, pages 50-56). Note the rat mean weight below is calculated from the pooled vehicle group, not a simple average of the two group means. The change from baseline in the table below is the simple difference between the means at the specified dates, and thus animals that die are only counted at the study initiation, not at the end of the study.

Table 7. Mean Weights and Changes (in g) in Rats

Dose Group	Dose	Males				Females			
		Week		Change from baseline	% change relative to vehicle	Week		Change from baseline	% change relative to vehicle
		-1	104			-1	104		
1. Vehicle	0	182.1	585.8	403.7		145.4	363.0	217.6	
2. Vehicle	0	183.0	595.4	412.4		144.8	374.4	229.6	
Pooled Vehicle	0	182.6	590.4	407.8		145.1	368.1	223.0	
Low	5	182.5	580.6	398.1	98%	145.9	365.6	219.7	98%
Medium	25	184.3	555.3	371.1	91%	144.4	344.6	200.2	90%
High	125	184.1	496.0	311.9	76%	144.2	325.2	181.0	81%

Table 8. Mean Weights and Changes (in g) in Mice

Dose Group	Males					Females				
	Dose	Week		Change from baseline	% change relative to vehicle	Dose	Week		Change from baseline	% change relative to vehicle
		-1	104				-1	104		
Vehicle	0	29.95	41.11	11.16		0	23.09	34.76	11.67	
Low	12.5	29.72	39.01	9.29	85%	25	22.64	32.80	10.16	87%
Medium	65	29.16	37.65	8.49	76%	100	22.94	32.05	9.11	78%
High	125	29.40	38.90	9.50	85%	300/200 /125	22.86	30.94	8.08	69%

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag *et al* (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' " From Table 5 and 6 above, the weight decrement criterion is quite exceeded in the high dose groups in rats and female mice (but note the initial dose in female mice is much higher than the corresponding dose in male mice). Although this requires the expertise of the toxicologist, this may be evidence that the MTD was exceeded in these dose groups. For rats, the Sponsor notes that "The high dose animals, (125 mg/kg/day) at 104 weeks had body weights 12 to 16% lower than control animal mean body weights. This is indicative that the highest dose, 125 mg/kg/day, the study exceeded the maximum tolerated dose (MTD)." (page 18 of rat report) In mice the Sponsor's report indicates that there were "statistically significant decreased mean body weights in all groups during the study." (page 23 of mouse report)

The Sponsor summarizes food consumption during the rat study as follows: "Group mean body weights were slightly decreased in male animals dosed 5 mg/kg/day [, i.e. the low dose,] and significantly decreased in both male and female rats dosed at ≥ 25.0 mg/kg/day when compared to the controls, there changes were considered to be treatment related." (pages 17-18 of rat report) In mice, the Sponsor reports that there were no consistent test article changes in food consumption.

Again from 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. This suggests that a useful way to assess whether or not the MTD was achieved is to measure early mortality not associated with any identified tumor. If this is high in the higher dose groups it suggests that animals tend to die before having time to develop tumors. Tables 9 and 10, below, displays the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors (i.e., the "Event"):

Table 9. Natural Death with No Identified Tumor in Rats (Male/Female)

	1. Vehicle	2.Vehicle	3. Low	4.Medium	5. High
Males Event	15	18	24	25	28
No event	45	42	36	35	32
Females Event	25	32	23	26	15
No event	35	28	37	34	45

It appears that in both rat genders there is evidence of a difference in the event distribution, but with different trends, increasing occurrences in males, somewhat decreasing in females. In both genders, time adjusted tests show differences in the event distribution across dose groups including the pooled vehicle (Males Overall: Logrank p = 0.0575, Wilcoxon p = 0.0434, Vehicle versus High: Logrank p = 0.0089, Wilcoxon p = 0.0077, Females Overall: Logrank p = 0.0389, Wilcoxon p = 0.0399, Vehicle versus High: Logrank p = 0.0414, Wilcoxon p = 0.0420). This may be evidence that the MTD was exceeded in male rats.

Table 10. Natural Death with No Identified Tumor in Mices (Male/Female)

	1. Vehicle	2. Low	3.Medium	4. High
Males Event	34	37	35	43
No event	26	23	25	17
Females Event	36	31	30	46
No event	24	29	30	14

There is weak evidence in males and strong evidence in females of early death without tumors (Males Overall: Logrank p = 0.0744, Wilcoxon p = 0.0237, Vehicle versus High: Logrank p = 0.0188, Wilcoxon p = 0.0055, Females Overall: both Logrank and Wilcoxon p $\leq$0.0001, Vehicle versus High: Logrank p = 0.0009, Wilcoxon p $\leq$0.0001). Once again, like the other observations above, these require the expertise of the toxicologist, but these tests provide evidence that the MTD was exceeded in both genders.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

This submission summarizes the results of two year rat and mouse studies to assess the carcinogenic potential carcinogenic potential of mifestrone following once daily dosing by gavage. The rat study was conducted [redacted] (b) (4), while the mouse study was conducted by [redacted] (b) (4)

2.2. Data Sources

The Sponsor provided two SAS transport files, each titled tumor.xpt containing a data set named tumor following the standard specifications of the FDA requested format. The submission for mice also included mortality, macro, micro, and signs transport files.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1. Study # 77389 Title: A 2-Year Oral Gavage Carcinogenicity Study of C-1073 in the Albino Rat.

STUDY DURATION: 104 Weeks (planned)
EXPERIMENTAL (DOSING) START DATE: 6 April 2004
TERMINAL SACRIFICE: Ended 21 April 2006
RAT STRAIN: Sprague Dawley Crl:CD®(SD) IGS BR Rats
ROUTE: Gavage

The rat study was conducted (b) (4).
The basic design of the rat study has three main drug dosing groups and two control groups as summarized in Table 10, below, actually a repeat of Table 1:

Table 11. Design of Albino Rat Study

Treatment Group	# Animals	Dosage (mg/kg/day)
1. Vehicle	60	0
2. Vehicle	60	0
3. Low	60	5
4. Medium	60	25
5. High	60	125

Treatment was applied by oral gavage with 0.25% carboxymethylcellulose and 0.2% Tween® 80 dissolved in sterile water as the vehicle. Dosing with the vehicle alone or test article was administered once daily at a dose volume of 10mL/kg/dose. In addition to these study groups, 10 animals of each gender were assigned to an undosed health screening group, and 11 animals of each gender were assigned to three toxicokinetic groups dosed at the levels of the low, medium and high dose groups. The Sponsor states that "Measured concentrations of C-1073 in the dose formulations deviated from nominal concentrations by a maximum of 5.1%" (page 36 of rat report). Animals were housed singly with water available ad libitum, with a diet of 4-5 pellets of a commercial laboratory diet offered daily.

The Sponsor states that "Seven days prior to treatment, all animals were weighed and randomly assigned to treatment group using a computer-based randomization procedure. Randomization was by stratification using body weight as the parameter. Males and females were randomized separately." (page 23 of rat report) The Sponsor reports that 10 animals that were considered to be in poor condition or dead were replaced in the first two weeks of the

study. This was not specified in the protocol, but the Sponsor claims should have no impact upon results.

"All animals were examined twice daily for mortality and signs of ill health or reaction to treatment. A complete detailed examination was performed weekly on main study animals. More frequent observations were undertaken if appropriate. A check for salivation was [usually] performed daily pre and post dosing on the high dose male animals. . . . In addition, from Week 26 onwards, all main study animals were examined for the presence of palpable masses during the detailed examination." (page 26 of rat report)

Dosing was justified as follows: "The dose levels were selected based on daily oral administration of C-1073 at dose levels of 1, 5, 25, and 125 mg/kg/day for 13 consecutive weeks to Sprague-Dawley rats, which resulted in significantly reduced body weight and body weight gain at 125 mg/kg/day . . ." (page 21 of report) A number of clinical pathology parameters were observed at the three higher doses, although the Sponsor note that: "As the findings were graded as minimal or slight, it was considered that 125 mg/kg/day would be suitable for the high dose in a 2 year study." (page 21 of report)

3.2.1.1. Sponsor's Results and Conclusions

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

Survival analysis:
The Sponsor provided the following summary of survival at the end of the 104-week treatment period:

Table 12. Sponsor Text Table 3: Survival

Group No.	Dose Level	Survival			
Identification	(mg/kg/day)	Males	%	Females	%
1/ Vehicle	0	44/60	73	32/60	53
2/ Vehicle	0	40/60	67	26/60	43
3/ C-1073	5.0	35/60	58	33/60	55
4/ C-1073	25.0	34/60	57	33/60	55
5/ C-1073	125.0	30/60	50	43/60	72

The report notes that: "An jncrease in the preterminal mortality of male rats given 125 mg/kg/day was seen." (page 36) But the Sponsor states that the reason for this increased mortality could not be determined from pathological investigations. Note there was no evidence of increased mortality in female rats.

Tumorigenicity analysis:

According to the Sponsor: "The administration of C-1073 did not produce the novo tumor types or promote rare tumor types, but rather exacerbated the development of spontaneous

thyroid and hepatic tumors commonly seen in the aging rat. Moreover, the administration of C-1073 was associated with decreased incidence of mammary gland tumor.

"Three thyroid follicular cell adenomas were seen in males given 125 mg/kg/day and eight thyroid follicular cell adenomas and three thyroid carcinomas occurred in females given 125 mg/kg/day. The incidences of follicular cell tumors are presented in [Sponsor's] Text Table 15, [see Table 12, below.] Thyroid follicular cell adenoma (males and females) and thyroid follicular cell carcinoma (females only) results in the high dose group were statistically significant and were considered to be an effect of C-1073 administration.

"A statistically significant increased incidence of hepatocellular adenoma was noted in females at ≥ 125 mg/kg/day. The incidence of hepatocellular adenoma is presented . . . [below.]

"The incidence of mammary gland fibroadenoma was decreased in a dose-dependent manner in treated females compared to both control groups and there was only one fibroadenoma recorded in high dose females." (page 49 of report) See below:

Table 13. Sponsor Text Table 15: Incidence of C-1073 Neoplastic Changes

Tissue/Finding	Sex	Male					Female				
Dose (mg/kg/day)		0	0	5	25	125	0	0	5	25	125
Thyroid	Number Examined	60	60	60	60	60	60	60	60	60	60
Adenoma: follicular cell											
	Total Number affected	1	0	0	1	3*	0	1	2	1	8*
Carcinoma: follicular cell											
	Total Number affected	0	1	0	1	0	0	0	1	0	3*
Liver	Number Examined	60	60	60	60	60	60	60	60	60	60
Adenoma: hepatocellular											
	Total Number affected	0	2	1	1	0	0	1	1	3	6*
Mammary Gland	Number Examined	50	48	54	56	54	59	60	59	59	60
Fibroadenoma											
	Total Number affected	2	2	1	3	1	22	17	10	6	1

* Statistically significant increase of tumor occurrence rate

The Sponsor claims that: "All other neoplastic changes that were seen in this study, including those reaching statistical significance, were generally typical of those commonly encountered in rats of this strain and age range." (page 49 of report).

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

The following tables (Table 14 for male rats, Table 15 for females) summarize the mortality results for the study groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. In these tables the terminal period only includes those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period. The Kaplan-Meier survival plots in Appendix 1 provide a more detailed picture of the profile of mortality losses.

Table 14. Summary of Male Rats Survival (dose at at 2 mg/cm^2)

Period (Weeks)	Vehicle 0	Vehicle 0	Low 10%	Medium 50%	High 100%
1-52	0/60[1]	0/60	3/60[1] 95.0%[2]	1/60 98.3%	3/60 95.0%
53-78	5/60 91.7%	6/60 90.0%	7/57 83.3%	9/59 83.3%	10/57 78.3%
79-91	4/55 85.0%	6/54 80.0%	7/50 71.7%	5/50 75.0%	6/47 68.3%
92-104	7/51 73.7%	8/48 66.7%	8/43 58.3%	11/45 56.7%	12/41 48,3%
Terminal [3] 105	44	40	35	34	29[4]

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

Table 15. Summary of Female Rat Survival (dose at 2 mg/cm^2)

Period (Weeks)	Vehicle 0	Vehicle 0	Low 10%	Medium 50%	High 100%
1-52	1/60[1] 98.3%[2]	1/60 98.3%	0/60	1/60 98.3%	2/60 96.7%
53-78	8/59 85.0%	11/59 80,0%	12/60 80.0%	10/59 81.7%	2/58 93.3%
79-91	9/51 70.0%	12/48 60.0%	8/48 66.7%	6/49 71.7%	5/56 85.0%
92-104	10/42 53.3%	10/36 43.3%	7/40 55.0%	10/43 55.0%	10/51 68.3%
Terminal [3] 105	32	26	33	33	41[4]

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

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

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1&2, 3-5	0.0206	0.0182	0.0815	0.0708
Homogeneity over Groups 1-5	0.0375	0.0345	0.0889	0.0838
No trend over Groups 1&2, 3-5	0.0081	0.0087	0.0129	0.0076
No Difference Between Groups 1&2 vs 5	0.0020	0.0019	0.0086	0.0067

Kaplan-Meier estimated survival curves across study groups for each gender are displayed in Figures A.1.1 and A.1.2 in Appendix 1. From Figure A.1.1, in male rats, at first there is a clear roughly decreasing survival over dose with the high dose group having the highest overall mortality. This is consistent with the results of the test above. For example, the tests of trend in mortality over the pooled vehicle groups with dosing groups 3-5 is highly statistically significant (logrank p = 0.0081, Wilcoxon p = 0.0087), confirmed by the pairwise test between the high dose and pooled controls (logrank p = 0.0020, Wilcoxon p = 0.0019). The tests of no homogeneity over all three dosing groups and pooled controls were statistically significant (logrank p = 0.0206, Wilcoxon p = 0.0182), as was the less appropriate test incorporating separate vehicle controls (logrank p = 0.0375, Wilcoxon p = 0.0345). In female rats results are quite inconsistent with these results. The high dose group has the highest survival, with vehicle group 2 having the lowest, and the other groups somewhat intertwined. This is sufficient to provide statistically significant evidence of a trend (logrank p = 0.0129, Wilcoxon p = 0.0076), but, again, the trend is for increasing survival over increasing dose. The difference between the high dose and pooled vehicle was also statistically significant (logrank p = 0.0086, Wilcoxon p = 0.0067). In female rats, the tests of no homogeneity over all five study

groups were not quite statistically significant (logrank p = 0.0889, Wilcoxon p = 0.0838), as were the more appropriate tests of no homogeneity using the pooled controls (logrank p = 0.0815, Wilcoxon p = 0.0708).

Results from a supporting experimental Bayesian nonparametric analysis of survival are provided in Appendix 2.

Tumorigenicity analysis:

As discussed in Section 1.3.1.4, for common tumors, the Haseman-Lin-Rahman rules for adjusting for multiplicity in a single species study specify that for a very rough 0.10 (10%) overall false positive error rate, both overall trend and the comparison between control and the high dose should be tested at a 0.05 (5%) level in rare tumors and at 0.01 (1%) level in common tumors.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.4, below, are often applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors (incidence > 1%) and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. As also discussed in section 1.3.1.4, using these adjustments for other tests, like the trend over the vehicle, low, and medium dose groups in mice and the pairwise comparisons between the vehicle and the medium-high, medium, and low dose groups can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate. The period '.' in these tables denotes the p-values of tests of dose groups with no tumors in any group.

Table 17. Potentially Statistically Significant Neoplasms in Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Trend	Significance Levels Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
Male									
HEMOLYM. TISSUE									
# Evaluated	60	60	60	60	60				
Malignant lymphoma	0	0	0	1	2	.0265	.0914	.3137	.
MAMMARY GLAND									
# Evaluated	50	48	54	56	54				
Adenoma/Adenocarcinoma	0	0	3	0	0	.7875	.	.	.0388
PITUITARY									
# Evaluated	60	60	60	60	60				
Adenoma: pars distalis	29	29	31	37	33	.1314	.1088	.0408	.2215
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Foll. cell	1	1	0	2	3	.0462	.1595	.3722	1
Adenoma: follicular cell	1	0	0	1	3	.0213	.0809	.5304	1

Table 17. (cont.) Potentially Statistically Significant Neoplasms in Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Trend	Significance Levels Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
Female									
KIDNEY									
# Evaluated	60	60	60	60	60				
Adenoma: tubular cell	0	0	0	0	2	.0480	.1320	.	.
LIVER									
# Evaluated	60	60	60	60	60				
Adenoma: hepatocellular	0	1	1	3	6	.0038	.0098	.1165	.5635
MAMMARY GLAND									
# Evaluated	59	60	59	59	60				
Adenocarcinoma	17	16	17	25	17	.7056	.6618	.0285	.4354
Adenoma	9	12	23	24	16	.5649	.1942	.0022	.0020
Adenoma/Adenocarcinoma	22	22	32	42	26	.6871	.3527	<0.0001	.0140
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Foll. cell	0	1	3	1	11	<0.0001	.0001	.5698	.1122
Adenoma: follicular cell	0	1	2	1	8	.0004	.0014	.5698	.2641
Carcinoma: follicular cell	0	0	1	0	3	.0187	.0471	.	.3381

Using the incidence in the pooled control group to specify whether a tumor is treated as common or rare, in both male and female rats, the test of trend in follicular cell adenoma of the thyroid is statistically significant (Males: $p = 0.0213$, Females: $p = 0.0004$, both < 0.025). In female rats the tests of trend in follicular cell carcinoma and pooled adenoma/carcinoma of the thyroid were both also statistically significant ($p = 0.0187$, $p < 0.0001$, respectively, < 0.025). Further, the pairwise comparisons between the pooled controls and high dose group in these three neoplasms in females were also statistically significant ($p = 0.0014$, 0.0471, $0.0001 < 0.05$). In female rats the test of trend in hepatocellular adenoma in the liver was also statistically significant ($p = 0.0038 < 0.025$) as was the pairwise comparison between the high dose group and pooled controls ($p = 0.0098 < 0.05$). Although this may extend the multiplicity correction into a region where one arguably should not go, the pairwise comparison between the medium dose groups versus the pooled controls in terms of pooled adenoma/adenocarcinoma of the mammary gland in female rats was also statistically significant ($p < 0.0001 < 0.01$). With the same caveat, the pairwise comparisons between controls and the medium and low dose groups in terms of mammary adenoma could also be classified as statistically significant ($p = 0.0022$ and $0.0020 < 0.01$, respectively). In male rats the test of trend in malignant lymphoma in hemolymphatic tissue was close to the multiplicity adjusted level of statistical significance ($p = 0.0265 \approx 0.025$). Finally, the count of three male rats in the low dose group with pooled adenoma and adenocarcinoma of the mammaries was sufficient to make the p-value of the test with the pooled controls small enough to be nominally significant in this extended criteria ($p = 0.0388 < 0.05$). No other comparisons reached the multiplicity adjusted statistical significance levels.

Complete incidence tables are provided in tables A.3.2 and A.3.3 of Appendix 3.

An alternative Bayesian analyses of carcinogenicity is presented in Appendix 4. Only tumors with sufficient animals and some possibility of a trend were analyzed. This analysis

suggests that there is no strong evidence of a dose related trend for any of the five tumors analyzed in males. However, in female rats the probability that the slope is non-zero in each of the tumors hepatocellular adenoma of the liver, follicular cell adenoma of the thyroid and pooled follicular cell adenoma and carcinoma of the thyroid is above 0.975. In fact the posterior probability of a dose related increase in the chance of developing a tumor is 0.99 or above. In female rats, note that 0 is near the middle of the remaining intervals for a linear dose response, again suggesting no strong evidence of a dose effect.

3.2.2. Research Study # 950-005 Title: 104-Week Oral Oncogenicity Study of C1073 in Mice.

STUDY DURATION: 104 Weeks (planned)
EXPERIMENTAL START DATE: 29 September 2004
TERMINAL SACRIFICE: 29 September 2006
MOUSE STRAIN: Charles River Crl:CD-1®(ICR) BR Mice
ROUTE: Gavage

Treatment was applied by oral gavage, with 0.25% carboxymethylcellulose and 0.2% Tween ® 80 dissolved in sterile water as a vehicle. These were dosed at a volume of 10mL/kg/dose. The mouse study was conducted (b) (4) . Gross aspects of the study designs for the main study animals are summarized below:

Table 18. Design of Mouse Study

Treatment Group	# Animals	Dosage (mg/kg/day)	
		Males	Females
1. Vehicle	60	0	0
2. Low	60	12.5	25
3. Medium	60	65	100
4. High	60	125	300/200/125[1]

[1]Females received 300 mg/kg from Weeks 1 to 35, 200 mg/kg from Weeks 36 to 53, and 125 mg/kg from Weeks 54 to 104.

The Sponsor noted that: "Mortality in females at 300 mg/kg was increased, resulting in the reduction of the dose level to 200 mg/kg at Week 36. Mortality in females at 300 [sic ,i.e. presumably meant 200,] remained increased, requiring a further reduction of the dose level to 125 mg/kg on study at Week 54. At 125 mg/kg, female survival was comparable to control." (page 22 of mouse report) In addition to the treatment groups noted above, for each of the three actual dose groups, 50 animals per gender were allocated for toxicokinetic analyses.

Animals were housed individually with food and water available *ad libitum*.

Dosing was justified as follows: "The dose levels were selected by the Sponsor on the basis of available data from (b) (4) Study Numbers 950-003 and 950-006. The high dose level of 125 mg/kg was selected for males based on a dose-related decrease in body weight gain

in males in [redacted] Study Numbers 950-003. Beginning on Week 5, and continuing until study termination, the males dosed at 65 and 125 mg/kg exhibited a dose-related decrease in mean body weight. At 125 mg/kg the decreases were statistically significant and ranged between 7 to 9% less than the control group. At 65 mg/kg, the decreases ranged between 4 to 7% less than the control group, but were only statistically significant in Weeks 6 and 7. Total mean body weight gain at 65 and 125 mg/kg was, respectively, 80 and 53% of the control group. The effect on body weight at 125 mg/kg was considered clearly related to treatment with C1073. The effect at 65 mg/kg was considered equivocal since the magnitude of body weight difference was low. The high dose of 300 mg/kg was selected for females based on the findings in both [redacted] Study Numbers 950-003 and 950-006. Females had dose-related increased incidence of cervical and vaginal diffuse squamous cell hyperplasia noted in [redacted] Study Number 950-003. The incidence and severity of the squamous cell hyperplasia, with all females at 125 mg/kg having these observations, was clearly dose related. In [redacted] Study Number 950-006, female animals were dosed at levels up to 1500 mg/kg/day for 28 consecutive days. There was mortality recorded in animals dosed 750, 1000, and 1500 mg/kg/day. Panlobular hepatocyte hypertrophy was noted in the liver of all animals dosed 500 mg/kg/day. Dose-related increases in both kidney and liver weights were observed in 125 through 750 mg/kg/day animals with the increases reaching statistical significance at 500 and 750 mg/kg/day. Cervical and vaginal diffuse squamous cell hyperplasia was noted in [redacted] Study Number 950-006 at dose levels above 250 mg/kg/day. Based on the findings in these studies, 300 mg/kg/day was selected as the high dose level for females in this study." (page 15 of mouse report)

3.2.1.1. Sponsor's Results and Conclusions

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

Survival analysis:

Note as discussed above, due to high mortality the dose level in the high dose group was reduced from 300 mg/kg to 200 mg/kg to 125 mg/kg. The Sponsor summarized mortality as follows: "Survival for males at all dose levels and females up to 100 mg/kg were comparable to controls during the first 18 months of study. . . . All groups remained until termination during [, i.e., some animals survived until] Week 104. Survival at the terminal necropsy of each male group was 25.0%, 23.3%, 25.0%, and 18.3% in the control, and 12.5, 65, and 125 mg/kg groups, respectively. Survival at the terminal necropsy of each female group was 23.3%, 33.3%, 31.7%, and 15.0% in the control, and 25, 100, and 300/200/125 mg/kg groups, respectively." (page 22 of report)

Tumorigenicity analysis:

According to the Sponsor, tumor incidence data were analyzed using both survival adjusted and unadjusted tests, specifically, Cochran-Armitage trend test, Fisher exact tests, and peto mortality adjusted tests. However, "There were no test article-related oncogenic effects in males or females. There were no statistically significant or biologically important differences in the incidences of neoplasms observed in treated and control males and females." (page 26 of report) Note the FDA analysis below is consistent with these conclusions.

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

The following tables (Table 19 for male mice, Table 20 for females) summarize the mortality results for the study groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. In these tables the terminal period only includes those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period. The Kaplan-Meier survival plots in Appendix 1 provide a more detailed picture of the profile of mortality losses. Note that the four animals excluded from the carcinogenicity analysis are included here (please see Section 2.2 for details.)

Table 19. Summary of Male Mice Survival (dose at at 2 mg/cm^2)

Period (Weeks)	Vehicle 0	Low 10%	Medium 50%	High 100%
1-52	13/60 78.3%	17/60 71.7%	13/60 78.3%	20/60 66.7%
53-78	12/47 58.3%	14/43 48.3%	15/47 53.3%	22/40 30.0%
79-91	14/35 35.0%	8/29 35.0%	11/32 35.0%	4/18 23.3%
92- 104	6/21 25.0%	8/21 21.7%	7/21 23.3%	3/14 18.3%
Terminal[3] 105	15	13	14	11[4]

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

Table 20. Summary of Female Mice Survival (dose at 2 mg/cm^2)

Period (Weeks)	Vehicle 0	Low 10%	Medium 50%	High 100%
1-52	10/60 83.3%	18/60 70.0%	9/60 85.0%	31/60 48.3%
53-78	14/50 60.0%	9/42 55.0%	19/51 53.3%	13/29 26.7%
79-91	9/36 45.0%	4/33 48.3%	9/32 38.3%	4/16 20.0%
92-104	13/27 23.3%	9/29 33.3%	4/23 31.7%	3/12 15.0%
Terminal[3] 105	14	20	19	9[4]

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

The following table, Table 20, summarizes the results from tests comparing survival profiles across study groups in the tumorigenicity data sets:

Table 21. Statistical Significances of Tests of Homogeneity and Trend in Survival

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Mice Homogeneity over Groups 1-4	0.1741	0.0482	0.0005	<0.0001
No trend over Groups 1-4	0.0677	0.0240	0.0002	<0.0001
No Difference Between Groups 1 vs 4	0.0364	0.0090	0.0021	0.0001

Figures A.1.3 and A.1.4, in Appendix 1, provide Kaplan-Meier dose group survival curves for study group within each mouse gender. From Figure A.1.3, in male mice, the high dose group has the highest overall mortality. The low dose group more or less has the next highest mortality, particularly in the middle of the study, with the vehicle and mid dose group mostly intertwined. This is consistent with the results of the tests above. For example, the tests of trend in mortality over study groups 1-4 has a statistically significant Wilcoxon statistic (p = 0.0240), while the logrank test is not quite significant at the commonly accepted 0.5 level (p = 0.0677). One might note that the log rank tests places greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus places more weight on earlier events than does the log rank test. There is statistically significant evidence of a difference between the high dose and vehicle (logrank p = 0.0364, Wilcoxon p = 0.0090). However, the somewhat inconsistent results in the outcomes of tests of survival differences continue when testing homogeneity in survival over all four groups in male mice (logrank p = 0.1741, Wilcoxon p = 0.0482). Results are somewhat different in female mice, especially in terms of statistical significance. Again, from Figure A.1.4, in female mice, the high dose group has the highest overall mortality, quite separated from the results in other study groups. Relatively speaking, survival in the other study groups tends to be somewhat intertwined, although again the low dose

group tends to have somewhat higher mortality in the middle of the study period. The much higher mortality in the high dose group is sufficient to explain the highly statistically significant tests for overall homogeneity, trend, and pairwise difference between the high dose and controls (all 3 logrank $p \leq 0.0021$, all 3 Wilcoxon $p \leq 0.0001$).

Results from a supporting experimental Bayesian nonparametric analysis of survival are provided in Appendix 2.

Tumorigenicity analysis:

As discussed in Section 1.3.1.4, for common tumors, the Haseman-Lin-Rahman rules for adjusting for multiplicity in a single species study specify that for a very rough 0.10 (10%) overall false positive error rate, both overall trend and the comparison between control and the high dose should be tested at a 0.05 (5%) level in rare tumors and at 0.01 (1%) level in common tumors.

Table 22. Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

	Incidence				Significance Levels			
Organ/ Tumor	Cntrl	Low	Med	Hi	Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
liver								
# Evaluated	60	60	60	60				
adenoma, hepatocellular[B]	1	0	5	2	.0334	.2637	.0914	1
lung								
# Evaluated	60	60	60	60				
Adenoma/Carcinoma Bronch. Alv.	10	7	9	10	.0370	.0857	.6783	.8188
carcinoma, bronch. alv.[M]	2	1	0	4	.0336	.0932	1	.8634

If we use the incidence in the control group no treatment group to specify whether a tumor is treated as common or rare, in both male and female mice no comparisons achieved even the loose levels of multiplicity adjusted significance. In males no tests even achieved the nominal 0.05 level.

Complete incidence tables in both species are provided in tables A.3.3 through A.3.6 below. An alternative Bayesian analysis presented in Appendix 4 leads to similar results.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see Section 1.1 above.

APPENDICES

Appendix 1. FDA Survival Analysis

Simple summary life tables in mortality are presented in the report (Tables 10, 11, 15, and 16, above). Kaplan-Meier estimated survival curves across study groups for each gender are displayed in Figures A.1.1 and A.1.2 for rats and Figures A.1.3 and A.1.4 for mice below. These plots include 95% confidence intervals around each survival curve (colored area around each curve). These plots are also supported by tests of homogeneity in survival over the different treatment groups including one test with the vehicle groups separate and one test with them pooled, tests of trend in survival over increasing dose over the groups, and the results of pairwise comparisons between the high dose group and the pooled control groups. The statistical significance levels (i.e., p-values) are provided in Tables A.1.1. and A.1.2. below. One might note that the log rank tests places greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus places more weight on earlier events than does the log rank test.

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rats Homogeneity over Groups 1-6	0.0206	0.0182	0.0815	0.0708
Homogeneity over Groups 1, 3-6	0.0375	0.0345	0.0889	0.0838
No trend over Groups 1, 3-6	0.0081	0.0087	0.0129	0.0076
No Difference Between Groups 1 vs 6	0.0020	0.0019	0.0086	0.0067

From Figure A.1.1, in male rats, at first there is a clear roughly decreasing survival over dose with the high dose group having the highest overall mortality. This is consistent with the results of the test above. For example, the tests of trend in mortality over the pooled vehicle groups with dosing groups 3-5 is highly statistically significant (logrank p = 0.0081, Wilcoxon p = 0.0087), confirmed by the pairwise test between the high dose and pooled controls (logrank p = 0.0020, Wilcoxon p = 0.0019). The tests of no homogeneity over all three dosing groups and pooled controls were statistically significant (logrank p = 0.0206, Wilcoxon p = 0.0182), as was the less appropriate test incorporating separate vehicle controls (logrank p = 0.0375, Wilcoxon p = 0.0345). In female rats results are quite inconsistent with these results. The high dose group has the highest survival, with vehicle group 2 having the lowest, and the other groups somewhat intertwined. This is sufficient to provide statistically significant evidence of a trend (logrank p = 0.0129, Wilcoxon p = 0.0076), but, again, the trend is for increasing survival over increasing dose. The difference between the high dose and pooled vehicle was also statistically significant (logrank p = 0.0086, Wilcoxon p = 0.0067). In female rats, the tests of no homogeneity over all five study groups were not quite statistically significant (logrank p = 0.0889, Wilcoxon p = 0.0838), as were the more appropriate tests of no homogeneity using the pooled controls (logrank p = 0.0815, Wilcoxon p = 0.0708).

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

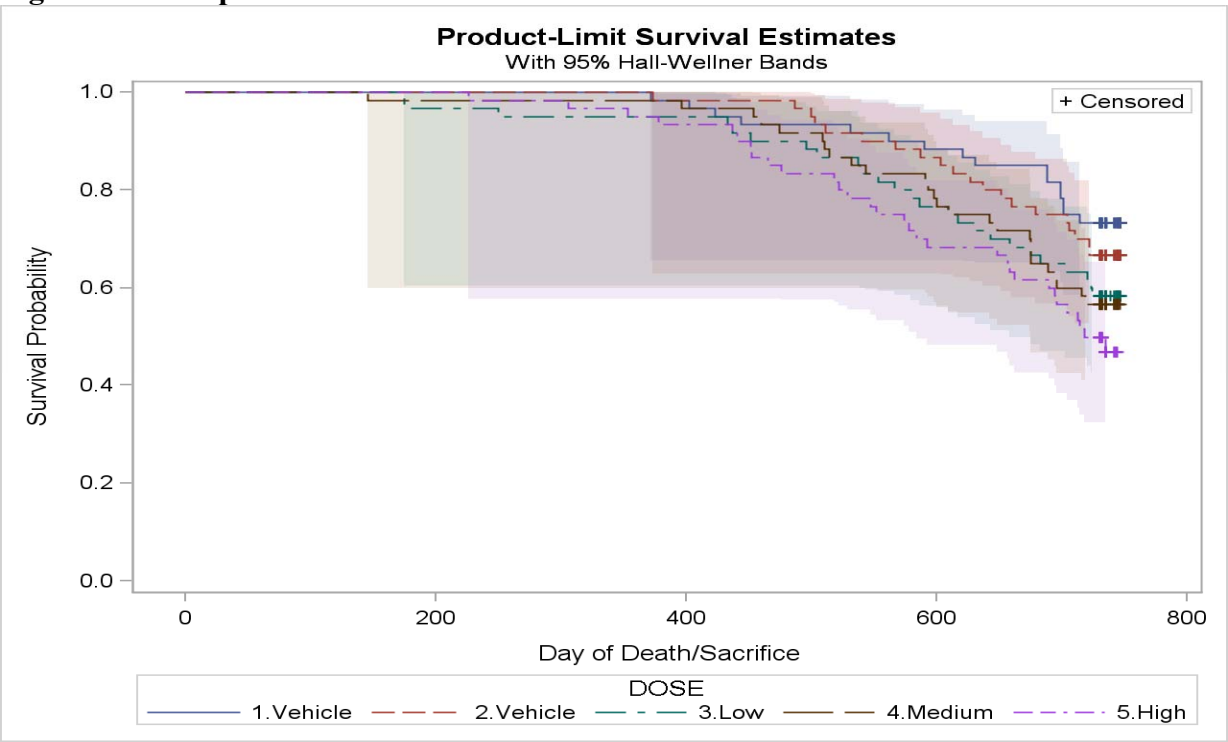


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats

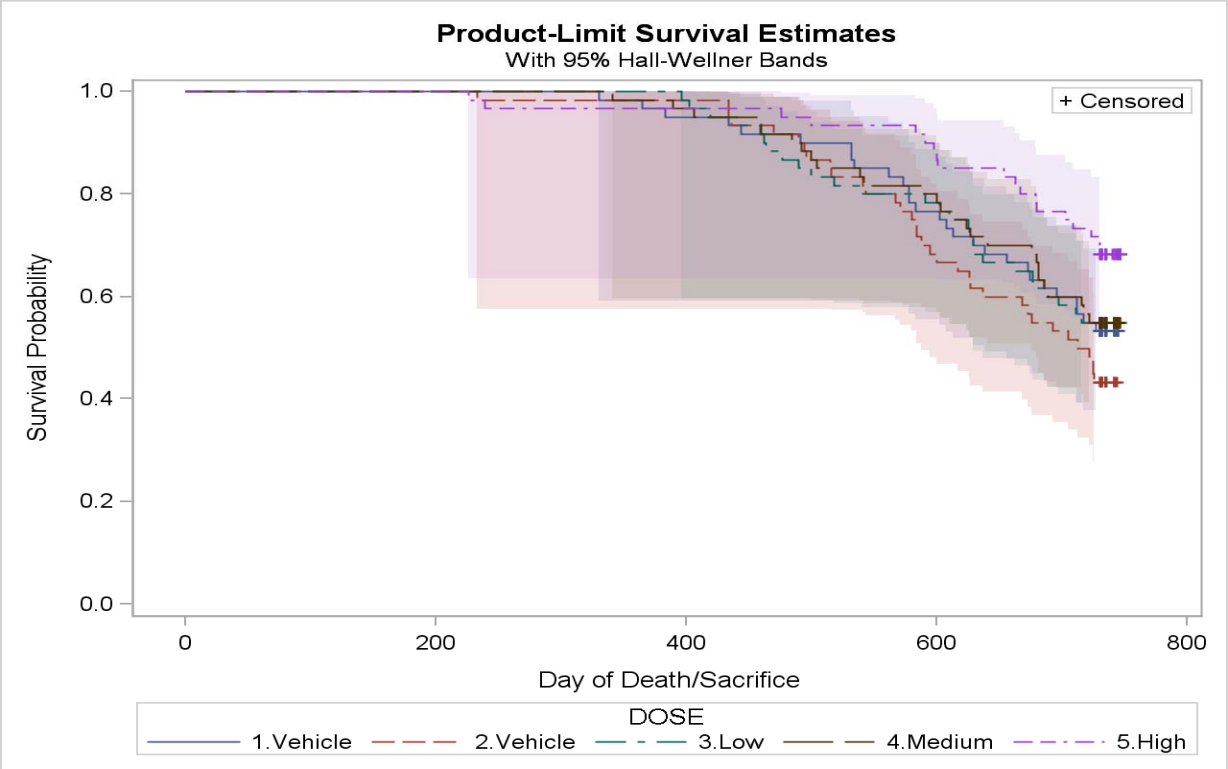


Table A.1.2. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Mouse Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Mouse Homogeneity over Groups 1-4	0.1741	0.0482	0.0005	<0.0001
No trend over Groups 1-4	0.0677	0.0240	0.0002	<0.0001
No Difference Between Groups 1 vs 4	0.0364	0.0090	0.0021	0.0001

Figures A.1.3 through A.1.4, below, provide similar survival curves for each mouse gender. From Figure A.1.3, in male mice, the high dose group has the highest overall mortality. The low dose group more or less has the next highest mortality, particularly in the middle of the study, with the vehicle and mid dose group mostly intertwined. This is consistent with the results of the tests above. For example the tests of trend in mortality over study groups 1-4 has a statistically significant Wilcoxon statistic ($p = 0.0240$), while the logrank test is not quite significant at the commonly accepted 0.5 level ($p = 0.0677$). As note in section 1.2.2 this is because the log rank test is more sensitive to later differences than early differences in survival. Note there is statistically significant evidence of a difference between the high dose and vehicle (logrank $p = 0.0364$, Wilcoxon $p = 0.0090$). The somewhat inconsistent results in the outcomes of tests of survival differences continue when testing homogeneity in survival over all four groups in male mice (logrank $p = 0.1741$, Wilcoxon $p = 0.0482$). Results are somewhat different in female mice, especially in terms of statistical significance. Again, from Figure A.1.4, in female mice, the high dose group has the highest overall mortality, quite separated from the results in other study groups. Relatively speaking, survival in the other study groups tends to be somewhat intertwined, although again the low dose group tends to have somewhat higher mortality in the middle of the study period. The much higher mortality in the high dose group is sufficient to explain the highly statistically significant tests for overall homogeneity, trend, and pairwise difference between the high dose and controls (all 3 logrank $p \leq 0.0021$, all 3 Wilcoxon $p \leq 0.0001$).

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

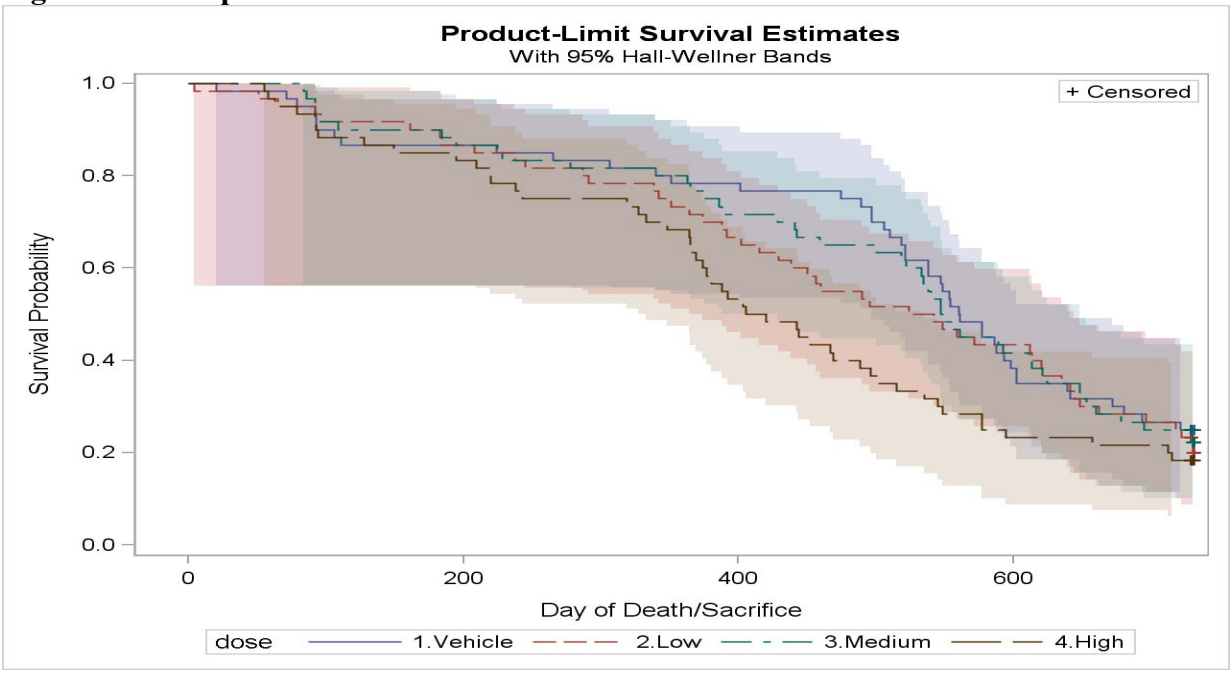
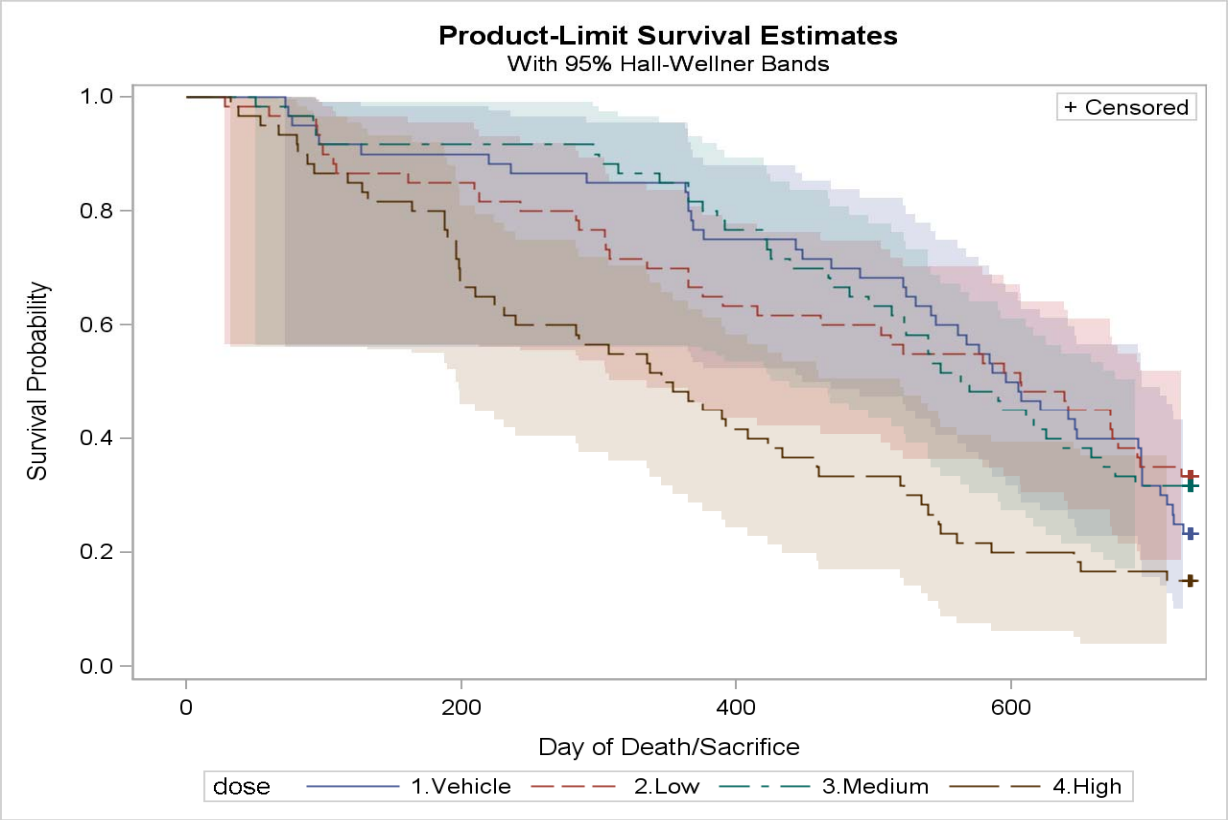


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



Appendix 2. FDA Nonparametric Bayesian Survival Analysis

The probability of a subject surviving past time t is given by the survival function, i.e., for random survival time T, S(t) = P(T > t). Statistical inference on survival is based on proposing a probability model for S(t) or one of its derivations. The probability model is defined so that hypotheses to be investigated are specified as parameters in the model. A frequentist analysis takes parameters as fixed and assesses the likelihood of the observed data. A Bayesian analysis starts by noting that parameters are not known, and assumes that a so-called prior probability distribution is a natural measure of this lack of exact knowledge. Then the Bayesian analysis assesses the impact of the actual observed data on this prior. In a nonparametric Bayesian analysis at least one of these parameters is the space of all probability distributions, or some large subset of this space. In other words, although a some prior weight is placed on a particular parametric family of distributions, the results would be consistent for other distributions. The actual nonparametric analysis used here is based upon using a so-called Dependent Dirichlet Process (DDP) as the prior on the space of all probability distributions.

Specifically, let T_i denote a random variable representing the survival time of the ith animal. For time until natural death time t_i we write $T_i = t_i$, but if the animal is sacrificed at time a_i, all we know is that the time until natural death is greater than a_i, written as $T_i \in (a_i, \infty)$, i.e. T_i is in the time interval (a_i, ∞). Note that animals whose death is in this interval are said to be censored. One useful probability model is to model the logarithm of T_i with a normal distribution, i.e., the T_i are modeled using a lognormal distribution. For this analysis, we model the distribution of $\log(T_i)$ as a mixture of normal distributions weighted by a Dirichlet process on the normal parameters. The prior is defined as a Dirichlet process where the baseline distribution models the linear parameters as a normal distribution on the linear mean parameters and the variance parameters with a Gamma distribution. The prior of the precision parameter of the Dirichlet process is a gamma distribution. The priors for the other hyperparameters are conjugate distributions. Mathematically we can write:

$$\log(T_i) = t_i \,|\, \mathrm{f}_{X_i} \sim \mathrm{f}_{X_i}$$
$$\mathrm{f}_{X_i} = \int N(X_i\beta, \sigma^2) G(d\beta\, d\sigma^2)$$
$$G \,|\, \alpha, G_0 \sim DP(\alpha G_0)$$

The distributions of the hyperparameters above are specified as follows:

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

This uses the LDDPsurvival function, for a Dirichlet Process mixture of normals, in the l DPpackage (Jara, 2007) of R (R Development Core Team, 2009). Currently, results should only be considered as supporting. The basic reference is de Iorio, et al (2009). The parameterization used to indicate doses was so-called dummy coding, which, in analogy with linear models as discussed in de Iorio et al (2004), implies that effect parameters for treatment doses correspond to the difference with the vehicle controls.

Thus the means mubd1 to mubd3 below indicate the differences between the pooled vehicle and the low, medium, and high dose groups (i.e., in rats. groups 3-5 versus pooled groups 1 & 2). The HPD interval is the estimated highest posterior density interval for the parameters. Conditional on the data, the probability the indicated parameter is in the HPD interval is 0.95. (As an aside, the numbers below are presented in the precision used in the program's output. It is to be expected that, at best, only the first or second digits are significant.)

Table A.2.1 Bayesian Parameter Estimates for Male Rats

Dummy Parameters	Mean	Median	Std. Dev.	95% HPD Lower	95% HPD Upper
mub(Intercept)	7.365430	7.293383	0.473380	6.508151	8.297804
mubd1	-0.030112	-0.014389	0.752438	-1.666670	1.362865
mubd2	0.254336	0.229497	0.699034	-1.228083	1.576904
mubd3	-0.004036	-0.024660	0.750328	-1.444792	1.590979
Precision parameters:					
tau2	0.489312	0.433447	0.274456	0.080992	1.012747
ncluster	32.835200	33.000000	8.203774	16.000000	47.000000

The highest posterior density (i.e., HPD) intervals corresponding to shortest intervals for the differences of vehicle group to each of the no treatment group and each treatment group whose posterior probability is 0.05. That is, conditioned on the data, they give the "best" estimates (in some sense) of the likely range of these differences. Note that 0 is near the middle of the intervals, indicating that there is no strong evidence of survival differences between the pooled vehicle group and any of the other study groups in males.

The tau2 and ncluster values are informative about the Dirichlet process used to derive these estimates and may be ignored.

The plots below show the estimated survival curves corresponding to the four actual doses within each species by gender. The survival curve of the control dose group is drawn as a solid line, the low dose as a dashed line, the medium dose as a dotted line, the high dose as a dot-dash line. Note that despite slight differences in survival curves there is no strong evidence of survival differences with the pooled controls.

Figure A.2.1 Bayesian Estimated Survival for Male Rats

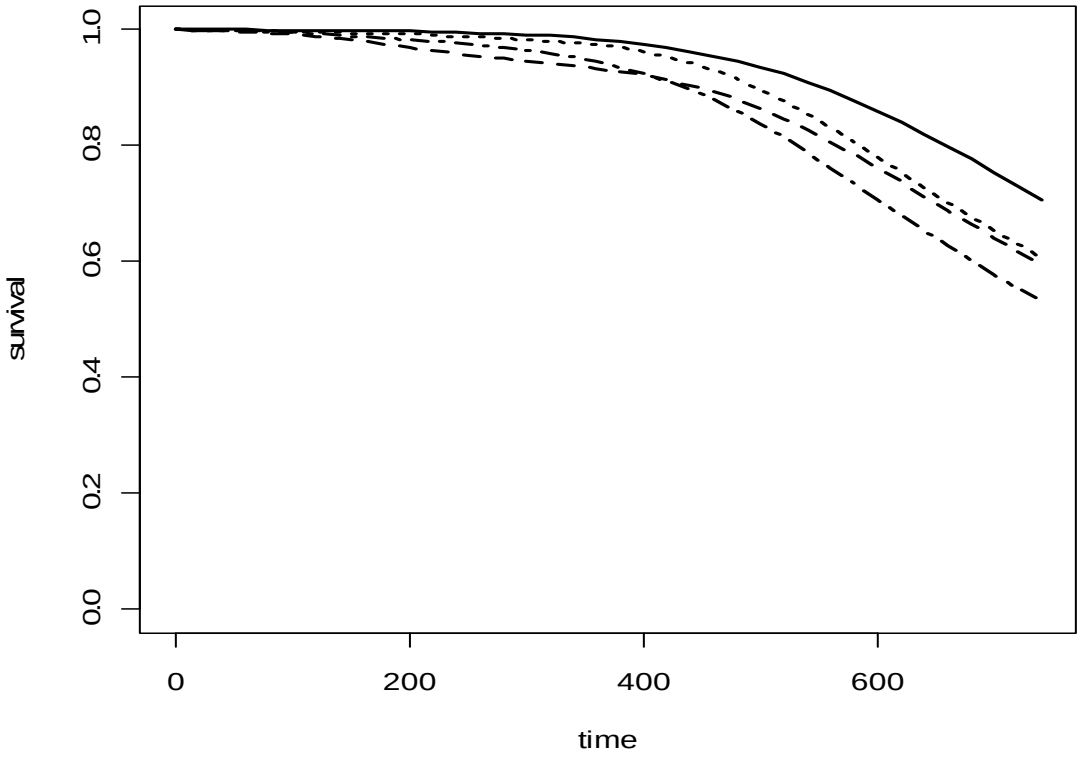
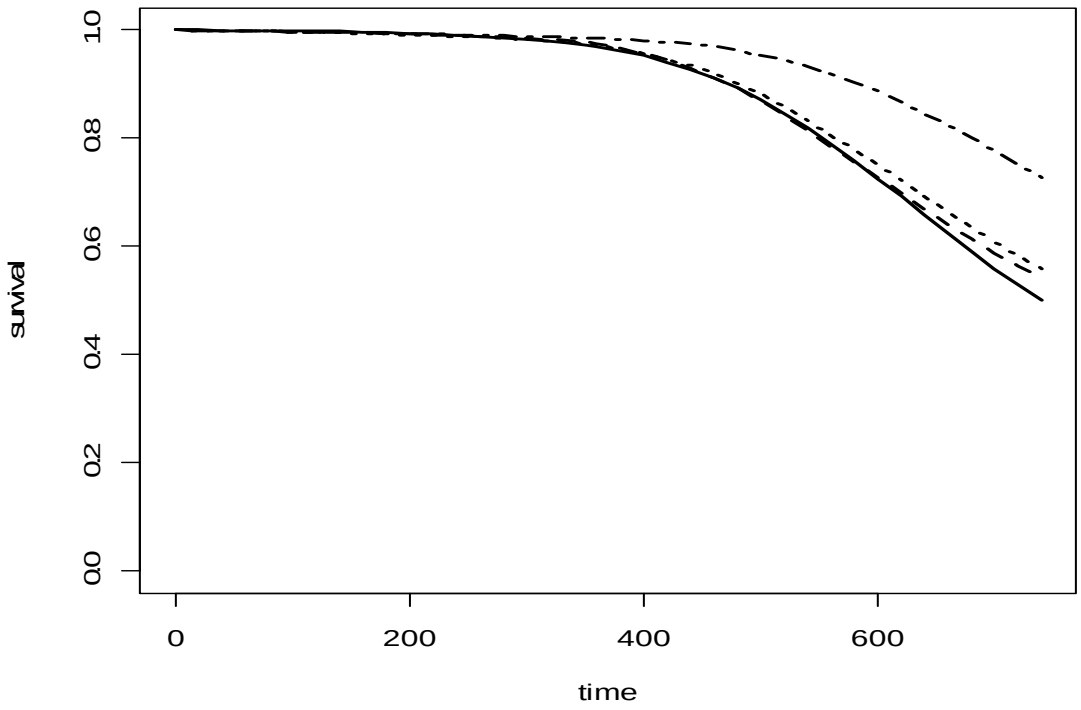


Table A.2.2 Bayesian Parameter Estimates for Female Rats

Dummy Parameters	Mean	Median	Std. Dev.	95% HPD Lower	95% HPD Upper
mub(Intercept)	6.934474	6.907492	0.414377	6.186339	7.784881
mubd1	0.707206	0.684820	0.799237	-0.842903	2.223073
mubd2	0.743642	0.689940	0.860103	-0.890842	2.492297
mubd3	0.806343	0.830302	0.806321	-0.704519	2.301205
Precision parameter:					
tau2	0.531244	0.468622	0.270479	0.125920	1.076342
ncluster	26.760800	27.000000	8.309596	9.000000	41.000000

Unlike males, there is weak evidence of greater survival in the dose groups than in the pooled vehicle controls. In particular, the probability that the AFT difference between the high dose and vehicle controls is positive is about 0.84. This is consistent with higher survival in each treatment group than in the pooled controls, as illustrated in Figure A.2.2 below.

Figure A.2.2 Bayesian Estimated Survival for Female Rats



Tables A.2.3 and A.2.4 and the corresponding figures are presented below.

Table A.2.3 Bayesian Parameter Estimates for Male Mice

Dummy Parameters	Mean	Median	Std. Dev.	95% HPD Lower	95% HPD Upper
mub(Intercept)	6.030246	6.038003	0.449494	5.140084	6.902905
mubd1	-0.258022	-0.248840	0.614678	-1.467080	0.975359
mubd2	0.102520	0.081536	0.589431	-1.020214	1.278192
mubd3	-0.097102	-0.113766	0.568273	-1.224608	0.995965
Precision parameter:					
tau2	0.378781	0.298502	0.287043	0.050442	0.951429
ncluster	35.647800	35.000000	7.073144	22.000000	49.000000

Again, 0 is near the middle of the intervals, indicating that there is no strong evidence of survival differences between the pooled vehicle group and any of the other study groups in male mice.

Figure A.2.3 Bayesian Estimated Survival for Male Mice

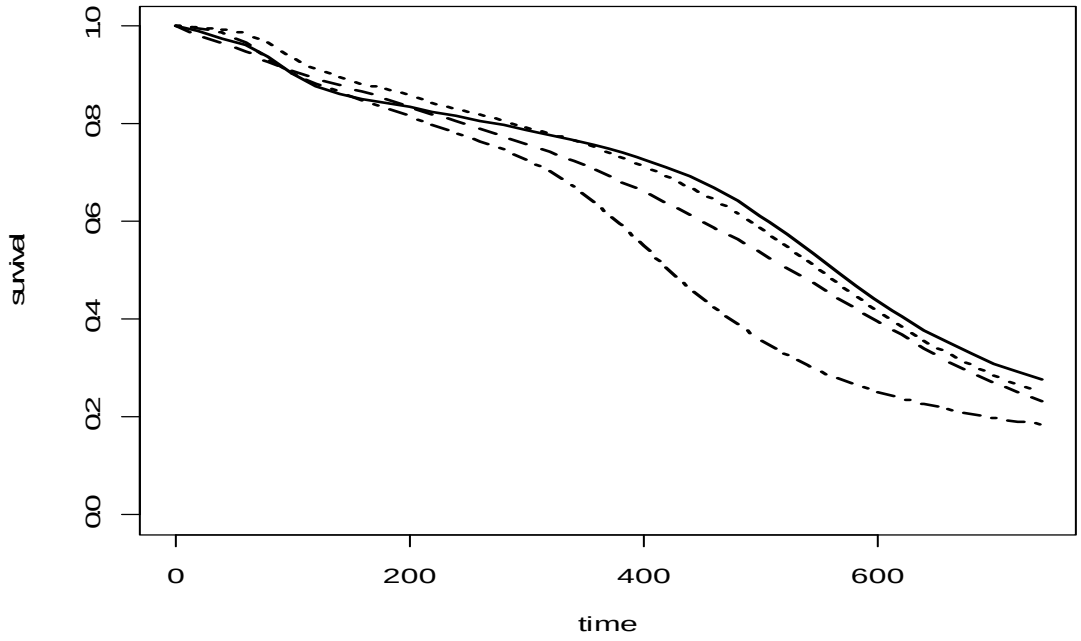
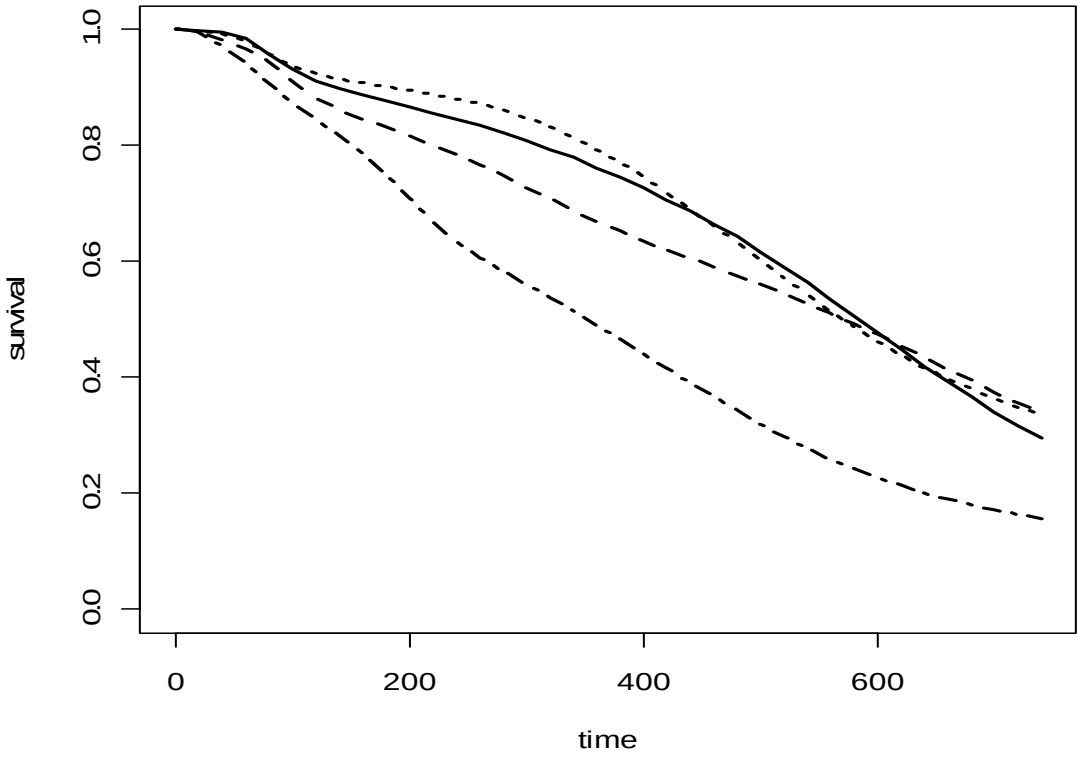


Table A.2.4 Bayesian Parameter Estimates for Female Mice

Dummy Parameters	Mean	Median	Std. Dev.	95% HPD Lower	95% HPD Upper
mub(Intercept)	6.154664	6.151133	0.331253	5.482526	6.799148
mubd1	-0.176010	-0.164961	0.470160	-1.137981	0.732167
mubd2	0.215929	0.210594	0.483433	-0.767203	1.131576
mubd3	-0.511217	-0.507913	0.475229	-1.498689	0.388728
Precision parameter:					
tau2	0.333237	0.288457	0.197338	0.074008	0.747569
ncluster	39.262200	39.000000	7.701358	23.000000	53.000000

The posterior probability that the difference between the high dose and the pooled vehicle is negative, corresponding to reduced survival in the high dose group, is roughly 0.84 or so. This is moderate, but not overwhelming evidence of differences.

Figure A.2.4 Bayesian Estimated Survival for Female Mice



Note that these analyses suggests no evidence of differences in survival in either gender. Since this is still an experimental approach, further research is required to investigate why the frequentist and Bayesian approaches suggest such different conclusions. It may simply be that the mixture of normals model used here is not appropriate for this data set, perhaps related to the fact that there are multiple crossings of survival curves.

Appendix 3. FDA Poly-k Tumorigenicity Analysis

The poly-k test, here with k=3, modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The tests used here are small sample exact permutation tests of tumor incidence. These do assume all marginal totals are fixed, a debatable assumption. This assumption implies that in the pairwise tests when one dose group has no tumors of the specific type and the other does, there is only one permutation of this pattern. Since that means that the only permutation of the data is the one observed, that means that all possible permutations are as extreme as the pattern observed, and thus the significance level of the observed pattern can be logically expressed as 1.0. One could use the same sort of argument when there were no tumors of the specific type being analyzed in either column of the 2x2 table corresponding to a pairwise comparison. Then an argument could be made that the p-value for this test should also be 1.0. However, largely for readability, in the tables below these p-values are considered as missing (i.e., corresponding to a null test), denoted by a period ".". Note that StatXact adjusts for the variance, which would be 0. Then the significance levels of the test statistics are based on the result of a division by 0, i.e., undefined, and hence StatXact codes these p-values as missing.

For each species by gender by organ the number of animals analyzed and used in the statistical tests is presented first. The tumor incidence for each organ is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high, medium, and low dose groups. These statistical tests are conditioned on the animals actually evaluated, ignoring those not analyzed. For tests in rats the two control groups are pooled.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.4 are often applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors (incidence > 1%) and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. As also discussed in section 1.3.1.4, using these adjustments for other tests, like the trend over the vehicle, low, and medium dose groups in mice and the pairwise comparisons between the vehicle and the medium-high, medium, and low dose groups can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Table A.3.1 in rats and Table A.3.2 in mice shows the tumors that had at least one mortality adjusted test whose nominal statistical significance was at least 0.05. Note that when one adjusts for multiplicity these nominally significant comparisons may not be statistically significant. Tables A.3.3 and A.3.4 display all incidences and statistical test results for male and female rats, respectively, while Tables A.3.5 and A.3.6 present similar results in male and female mice. The p-values of the poly-k test are based on exact tests from StatXact as discussed above. As also noted above, the period '.' denotes the p-values of tests of dose groups with no tumors in any group.

Table A.3.1 Potentially Statistically Significant Neoplasms in Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Trend	Significance Levels Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
Male									
HEMOLYM. TISSUE									
# Evaluated	60	60	60	60	60				
Malignant lymphoma	0	0	0	1	2	.0265	.0914	.3137	.
MAMMARY GLAND									
# Evaluated	50	48	54	56	54				
Adenoma/Adenocarcinoma	0	0	3	0	0	.7875	.	.	.0388
PITUITARY									
# Evaluated	60	60	60	60	60				
Adenoma: pars distalis	29	29	31	37	33	.1314	.1088	.0408	.2215
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Foll. cell	1	1	0	2	3	.0462	.1595	.3722	1
Adenoma: follicular cell	1	0	0	1	3	.0213	.0809	.5304	1
Female									
KIDNEY									
# Evaluated	60	60	60	60	60				
Adenoma: tubular cell	0	0	0	0	2	.0480	.1320	.	.
LIVER									
# Evaluated	60	60	60	60	60				
Adenoma: hepatocellular	0	1	1	3	6	.0038	.0098	.1165	.5635
MAMMARY GLAND									
# Evaluated	59	60	59	59	60				
Adenocarcinoma	17	16	17	25	17	.7056	.6618	.0285	.4354
Adenoma	9	12	23	24	16	.5649	.1942	.0022	.0020
Adenoma/Adenocarcinoma	22	22	32	42	26	.6871	.3527	<0.0001	.0140
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Foll. cell	0	1	3	1	11	<0.0001	.0001	.5698	.1122
Adenoma: follicular cell	0	1	2	1	8	.0004	.0014	.5698	.2641
Carcinoma: follicular cell	0	0	1	0	3	.0187	.0471	.	.3381

Using the incidence in the pooled control group to specify whether a tumor is treated as common or rare (i.e., more or less than 1%), in both male and female rats, the test of trend in follicular cell adenoma of the thyroid is statistically significant (Males: $p = 0.0213$, Females: $p = 0.0004$, both < 0.025). In female rats the tests of trend in follicular cell carcinoma and pooled adenoma/carcinoma of the thyroid were both also statistically significant ($p = 0.0187$, $p < 0.0001$, respectively, both < 0.025). Further, the pairwise comparisons between the pooled controls and high dose group in these three neoplasms in females were also statistically significant ($p = 0.0014, 0.0471, 0.0001 < 0.05$). In female rats the test of trend in hepatocellular adenoma in the liver was also statistically significant ($p = 0.0038 < 0.025$) as was the pairwise comparison between the high dose group and pooled controls ($p = 0.0098 < 0.05$). Although this may well extend the multiplicity correction into a region where one arguably should not go, the pairwise comparison between the medium dose groups versus the pooled controls in terms of pooled adenoma/adenocarcinoma of the mammary gland in female rats was also statistically significant ($p < 0.0001 < 0.01$). With the same caveat, the pairwise comparisons between controls and the medium and low dose groups in terms of mammary adenoma could also be classified as statistically significant ($p = 0.0022$ and $0.0020 < 0.01$, respectively). In male rats

the test of trend in malignant lymphoma in hemolymphatic tissue was close to the multiplicity adjusted level of statistical significance ($p = 0.0265 \approx 0.025$). Finally, the count of three male rats in the low dose group with pooled adenoma and adenocarcinoma of the mammaries was sufficient to make the p-value of the test with the pooled controls small enough to be nominally significant in this extended criteria ($p = 0.0388 < 0.05$). No other comparisons reached the multiplicity adjusted statistical significance levels.

Table A.3.2 Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

	Incidence				Significance Levels			
Organ/ Tumor	Cntr1	Low	Med	Hi	Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
liver								
# Evaluated	60	60	60	60				
adenoma, hepatocellular[B]	1	0	5	2	.0334	.2637	.0914	1
lung								
# Evaluated	60	60	60	60				
Adenoma/Carcinoma Bronch. Alv.	10	7	9	10	.0370	.0857	.6783	.8188
carcinoma, bronch. alv.[M]	2	1	0	4	.0336	.0932	1	.8634

If we use the incidence in the control group no treatment group to specify whether a tumor is treated as common or rare, in both male and female mice no comparisons achieved even the loose levels of multiplicity adjusted significance. In males not tests even achieved the nominal 0.05 level.

Complete incidence tables in rats are provided in tables A.3.3 and A.3.4 below.

Table A.3.3 Incidence and Significance Levels of all Tests on Neoplasms in Male Rats

	Incidence					Significance Levels			
Organ/ Tumor	C1	C2	Low	Md	Hi	Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
ABDOMEN									
# Evaluated	2	0	3	1	0				
Lipoma	0	0	0	1	0	.3333	.	.5000	.
Malignant schwannoma	0	0	2	0	0	.3333	.	.	.3333
Osteosarcoma	1	0	0	0	0	1	.	.	1
ADRENAL									
# Evaluated	60	60	60	60	60				
Benign pheochromocytoma	3	1	2	1	0	.9358	1	.8549	.5990
Malignant pheochromocytoma	1	2	1	0	0	.9666	1	1	.7737
Pheochromocytoma [B]&[M]	4	3	3	1	0	.9895	1	.9561	.6429
BRAIN									
# Evaluated	60	60	60	60	60				
Benign granular cell tumor	0	0	0	1	0	.3796	.	.3137	.
Benign oligodendroglioma	1	0	0	0	0	1	1	1	1
Malignant astrocytoma	1	2	1	1	2	.2831	.4649	.7769	.7769
Malignant meningioma	1	0	0	0	0	1	1	1	1
Papilloma: choroid plexus	0	0	1	0	0	.5732	.	.	.3137
CAVITY CRANIAL									
# Evaluated	0	1	0	0	1				
Malignant schwannoma	0	0	0	0	1	1	.	.	.
CAVITY ORAL									
# Evaluated	0	1	1	0	0				
Carcinoma: squamous cell	0	1	0	0	0	1	.	.	.

Table A.3.3 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Male Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Significance Levels Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
CECUM									
# Evaluated	60	60	60	60	60				
Adenocarcinoma	0	0	0	1	0	.3796	.	.3137	.
EYE									
# Evaluated	60	60	59	60	60				
Benign melanoma: uvea	0	1	0	0	0	1	1	1	1
HEMOLYM. TISSUE									
# Evaluated	60	60	60	60	60				
Histiocytic sarcoma	0	1	2	1	1	.3674	.5114	.5304	.2319
Leukemia: large granular lymphoc	1	0	0	0	1	.3330	.5086	1	1
Malignant lymphoma	0	0	0	1	2	.0265	.0914	.3137	.
Mast cell tumor	0	0	1	0	0	.5714	.	.	.3092
JEJUNUM									
# Evaluated	60	60	60	60	60				
Adenocarcinoma	0	1	0	0	0	1	1	1	1
Leiomyosarcoma	1	0	0	0	0	1	1	1	1
KIDNEY									
# Evaluated	60	60	60	60	60				
Adenoma: tubular cell	0	0	0	0	1	.1837	.3000	.	.
Carcinoma: transitional cell	0	0	0	0	1	.1837	.3000	.	.
LACRIMAL GLAND									
# Evaluated	60	60	60	60	60				
Fibroma	0	0	1	0	0	.5714	.	.	.3092
LIVER									
# Evaluated	60	60	60	60	60				
Adenoma: hepatocellular	0	2	1	1	0	.7926	1	.6797	.6733
MAMMARY GLAND									
# Evaluated	50	48	54	56	54				
Adenocarcinoma	0	0	2	0	0	.6759	.	.	.1163
Adenoma	0	0	1	0	0	.6000	.	.	.3385
Adenoma/Adenocarcinoma	0	0	3	0	0	.7875	.	.	.0388
Fibroadenoma	2	2	1	3	1	.6321	.8603	.4472	.8758
PANCREAS									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; islet cell	2	1	5	4	2	.5278	.4732	.1393	.0641
Adenoma: islet cell	2	1	4	2	2	.4504	.4732	.5015	.1393
Carcinoma: acinar cell	1	0	0	0	0	1	1	1	1
Carcinoma: islet cell	0	0	1	2	0	.4866	.	.0970	.3092
PARATHYROID GLAND									
# Evaluated	52	51	47	53	51				
Adenoma	0	1	0	0	1	.3248	.4962	1	1
PITUITARY									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; pars distalis	29	31	33	38	33	.2034	.1538	.0535	.1538
Adenoma: pars distalis	29	29	31	37	33	.1314	.1088	.0408	.2215
Carcinoma: pars distalis	0	2	2	1	0	.8175	1	.6797	.3722
Pituicytoma: pars nervosa	0	0	1	0	0	.5714	.	.	.3092
RECTUM									
# Evaluated	60	60	60	60	60				
Adenoma	0	0	0	1	0	.3796	.	.3137	.

Table A.3.3 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Male Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Significance Levels Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
SKIN MISCELLANEOUS									
# Evaluated	21	27	15	15	14				
Adenoma: basal cell	1	0	0	0	0	1	1	1	1
Carcinoma: squamous cell	0	1	0	0	0	1	1	1	1
Hemangioma	0	1	0	0	0	1	1	1	1
Keratoacanthoma	0	0	0	1	0	.2949	.	.2321	.
Papilloma/carcinoma; Sq. cell	2	1	1	3	0	.6582	1	.1246	.6304
Papilloma: squamous cell	2	0	1	3	0	.5938	1	.0760	.5296
SPINAL CORD CERVIC									
# Evaluated	60	60	60	60	60				
Malignant astrocytoma	0	0	0	1	0	.3821	.	.3182	.
STOMACH									
# Evaluated	60	60	60	60	60				
Leiomyosarcoma	0	1	1	0	0	.8153	1	1	.5214
SUBCUTANEOUS TISSU									
# Evaluated	8	8	8	9	7				
Fibroma	0	2	3	3	0	.7871	1	.2076	.1603
Fibrosarcoma	1	2	1	2	1	.4567	.7183	.5862	.8134
Hemangiosarcoma	0	1	0	1	0	.6118	1	.5850	1
Malignant fibrous histiocytoma	1	0	0	0	0	1	1	1	1
Systemic									
# Evaluated	60	60	60	60	60				
Hemangioma/-Sarcoma	0	2	0	1	1	.3799	.6570	.6768	1
TAIL									
# Evaluated	3	2	5	3	4				
Hemangioma	0	0	0	0	1	.2308	.4286	.	.
Papilloma: squamous cell	2	0	1	0	1	.6895	.8857	1	.9286
TESTIS									
# Evaluated	60	60	59	60	60				
Adenoma/Carcinoma; Intrst. cell	3	2	2	1	2	.5158	.6721	.8984	.6914
Adenoma: interstitial cell	2	2	1	1	2	.3773	.5791	.8501	.8451
Carcinoma: interstitial cell	1	0	1	0	0	.8173	1	1	.5242
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; C-cell	2	1	1	0	2	.2853	.4690	1	.7737
Adenoma/Carcinoma; Foll. cell	1	1	0	2	3	.0462	.1595	.3722	1
Adenoma: C-cell	2	1	1	0	2	.2853	.4690	1	.7737
Adenoma: follicular cell	1	0	0	1	3	.0213	.0809	.5304	1
Carcinoma: follicular cell	0	1	0	1	0	.6161	1	.5304	1
URINARY BLADDER									
# Evaluated	60	60	59	60	58				
Carcinoma: transitional cell	0	0	0	0	1	.1811	.2953	.	.

Table A.3.4 Incidence and Significance Levels of all Tests on Neoplasms in Female Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Significance Levels Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
ABDOMEN									
# Evaluated	1	2	0	1	0				
Paraganglioma (M)	1	0	0	0	0	1	.	.	.
ADRENAL									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; Cortical	0	1	1	2	1	.3778	.5990	.2705	.5635
Adenoma: cortical	0	1	1	0	1	.4311	.5990	1	.5635
Benign pheochromocytoma	1	2	1	1	1	.6662	.8418	.8177	.8124
Carcinoma: cortical	0	0	0	2	0	.4329	.	.1159	.
Malignant pheochromocytoma	0	1	0	0	0	1	1	1	1
Pheochromocytoma [B]&[M]	1	3	1	1	1	.7580	.9013	.8821	.8777
BRAIN									
# Evaluated	60	60	60	60	60				
Malignant ependymoma	0	1	0	0	0	1	1	1	1
CAVITY ORAL									
# Evaluated	0	0	1	0	0				
Carcinoma: squamous cell	0	0	1	0	0	1	.	.	.
CLITORAL GLAND									
# Evaluated	1	0	0	0	1				
Adenoma	1	0	0	0	0	1	1	.	.
DUODENUM									
# Evaluated	60	60	60	60	60				
Adenoma	0	0	1	0	0	.6167	.	.	.3381
HEMOLYM. TISSUE									
# Evaluated	60	60	60	60	60				
Histiocytic sarcoma	0	0	1	0	0	.6167	.	.	.3381
Malignant lymphoma	0	0	0	1	1	.1368	.3655	.3429	.
KIDNEY									
# Evaluated	60	60	60	60	60				
Adenoma: tubular cell	0	0	0	0	2	.0480	.1320	.	.
Carcinoma: tubular cell	0	0	0	1	0	.4208	.	.3429	.
LIVER									
# Evaluated	60	60	60	60	60				
Adenoma: hepatocellular	0	1	1	3	6	.0038	.0098	.1165	.5635
MAMMARY GLAND									
# Evaluated	59	60	59	59	60				
Adenocarcinoma	17	16	17	25	17	.7056	.6618	.0285	.4354
Adenoma	9	12	23	24	16	.5649	.1942	.0022	.0020
Adenoma/Adenocarcinoma	22	22	32	42	26	.6871	.3527	<0.0001	.0140
Fibroadenoma	22	17	10	6	1	1	1	.9999	.9938
MUSCLE SKELETAL MI									
# Evaluated	0	0	1	0	0				
Hemangiosarcoma	0	0	1	0	0	1	.	.	.
OVARY									
# Evaluated	60	60	60	60	60				
Benign thecoma	1	0	0	0	0	1	1	1	1
PANCREAS									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; islet cell	3	3	2	1	3	.4898	.6994	.9499	.8168
Adenoma: acinar cell	0	1	0	0	0	1	1	1	1
Adenoma: islet cell	3	3	0	0	2	.5607	.8566	1	1
Carcinoma: islet cell	0	0	2	1	1	.3013	.3655	.3429	.1127

Table A.3.4 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Female Rats

Organ/ Tumor	Incidence C1	C2	Low	Md	Hi	Significance Levels Trend	Hi vs C1+C2	Med vs C1+C2	Low vs C1+C2
PARATHYROID GLAND									
# Evaluated	42	48	50	46	53				
Carcinoma	0	0	0	0	1	.2474	.4123	.	.
PITUITARY									
# Evaluated	58	60	60	60	60				
Adenoma/Carcinoma; pars distalis	51	52	57	56	52	.6786	.6051	.1844	.1761
Adenoma: pars distalis	46	43	49	49	48	.4665	.4263	.3210	.2401
Carcinoma: pars distalis	5	9	8	7	4	.9255	.9434	.6192	.4899
SKIN MISCELLANEOUS									
# Evaluated	24	15	18	19	14				
Adenoma: basal cell	0	0	1	0	0	.5584	.	.	.3061
Benign schwannoma	0	0	0	0	1	.1558	.2609	.	.
Carcinoma: squamous cell	0	0	1	0	0	.5584	.	.	.3061
Leiomyoma	0	0	0	0	1	.1558	.2609	.	.
Papilloma/carcinoma; Sq. cell	0	1	1	1	0	.6254	1	.5420	.5230
Papilloma: squamous cell	0	1	0	1	0	.5981	1	.5420	1
STOMACH									
# Evaluated	60	60	60	60	60				
Leiomyoma	1	0	0	0	0	1	1	1	1
Leiomyosarcoma	1	0	0	0	0	1	1	1	1
Papilloma: squamous cell	0	1	0	0	0	1	1	1	1
SUBCUTANEOUS TISSU									
# Evaluated	2	4	2	0	3				
Fibroma	0	1	1	0	1	.2857	.5238	.	.3333
Lipoma	1	0	0	0	0	1	1	.	1
Malignant schwannoma	0	0	1	0	1	.1071	.2857	.	.1667
Systemic									
# Evaluated	60	60	60	60	60				
Hemangioma/-Sarcoma	0	0	1	0	0	.6167	.	.	.3381
TAIL									
# Evaluated	4	1	5	3	3				
Papilloma: squamous cell	1	0	0	0	0	1	1	1	1
THYROID									
# Evaluated	60	60	60	60	60				
Adenoma/Carcinoma; C-cell	1	0	3	1	0	.8358	1	.5698	.1165
Adenoma/Carcinoma; Foll. cell	0	1	3	1	11	<0.0001	.0001	.5698	.1122
Adenoma: C-cell	1	0	2	1	0	.7748	1	.5698	.2641
Adenoma: follicular cell	0	1	2	1	8	.0004	.0014	.5698	.2641
Carcinoma: C-cell	0	0	1	0	0	.6167	.	.	.3381
Carcinoma: follicular cell	0	0	1	0	3	.0187	.0471	.	.3381
TONGUE									
# Evaluated	60	60	60	60	60				
Benign granular cell tumor	0	0	0	0	1	.2208	.3655	.	.
UTERUS									
# Evaluated	59	60	60	60	60				
Benign granular cell tumor	1	1	0	0	0	1	1	1	1
Polyp: endometrial stromal	5	5	1	0	0	1	1	1	.9918
Sarcoma/Polyp; endo. stromal	7	5	1	0	3	.7797	.9600	1	.9966
Sarcoma: endometrial stromal	2	0	0	0	3	.0731	.2567	1	1
VAGINA									
# Evaluated	59	60	60	60	60				
Benign granular cell tumor	2	0	0	0	0	1	1	1	1

Complete incidence tables in mices are provided in tables A.3.5 and A.3.6 below.

Table A.3.5 Incidence and Significance Levels of all Tests on Neoplasms in Male Mice

Organ/ Tumor	Incidence Cntr1	Low	Med	Hi	Significance Levels Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
adrenal glands								
# Evaluated	58	55	56	58				
adenoma, subcapsular cell[B]	2	0	0	0	1	1	1	1
pheochromocytoma[B]	0	1	0	0	.7041	.	.	.4423
cavity, thoracic								
# Evaluated	60	55	56	60				
osteosarcoma[M]	0	1	0	0	.6931	.	.	.4364
epididymides								
# Evaluated	60	55	56	59				
schwannoma[M]	0	0	1	0	.4653	.	.4561	.
harderian glands								
# Evaluated	60	55	55	60				
adenoma[B]	4	1	2	2	.5834	.8051	.8465	.9463
kidneys								
# Evaluated	60	55	56	60				
adenoma, tubular cell[B]	1	0	0	0	1	1	1	1
nephroblastoma[B]	0	0	1	0	.4600	.	.4464	.
liver								
# Evaluated	60	55	56	60				
Adenoma/Carcinoma; hepato.	5	2	6	2	.5162	.8445	.3284	.8717
adenoma, hepatocellular[B]	2	2	5	1	.4356	.7966	.1452	.5736
carcinoma, hepatocellular[M]	3	0	1	1	.6587	.8772	.9090	1
osteosarcoma[M]	0	1	0	0	.6931	.	.	.4364
lung								
# Evaluated	60	55	56	60				
Adenoma/Carc. Bronch. Alv.	12	11	6	8	.7263	.6243	.9315	.3867
adenoma, bronch. alv.[B]	12	10	5	7	.8094	.7465	.9607	.5059
carcinoma, bronch. alv.[M]	0	1	1	1	.2251	.4038	.4561	.4259
osteosarcoma[M]	0	1	0	0	.6931	.	.	.4364
sarcoma, undiff.[M]	1	1	1	0	.7786	1	.7086	.6751
multicentric neoplasm								
# Evaluated	60	55	56	60				
Hemagiosarcoma/Hemangioma	6	2	1	1	.9542	.9759	.9858	.9295
hemangioma[B]	1	0	0	0	1	1	1	1
hemangiosarcoma[M]	5	2	1	1	.9263	.9605	.9750	.8917
lymphoma[M]	3	3	1	2	.6430	.6804	.9203	.5344
sarcoma, histiocytic[M]	2	2	0	0	.9635	1	1	.5780
pancreas								
# Evaluated	59	55	56	60				
adenoma, islet cell[B]	0	0	1	1	.1515	.4118	.4545	.
pharynx								
# Evaluated	60	54	55	59				
papilloma, squamous cell[B]	0	0	0	1	.2100	.4038	.	.
pituitary gland								
# Evaluated	55	53	54	58				
Adenoma Pars dist./inter.	0	0	0	1	.2105	.4082	.	.
adenoma, pars distalis[B]	0	0	0	1	.2105	.4082	.	.
skin								
# Evaluated	60	55	56	60				
carcinoma, squamous cell[M]	0	0	0	1	.2100	.4038	.	.

Table A.3.5 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Male Mice

Organ/ Tumor	Incidence Cntr1	Low	Med	Hi	Significance Levels Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
skin, subcutis								
# Evaluated	60	55	56	60				
fibrous histiocytoma[M]	0	1	0	0	.6900	.	.	.4259
sarcoma, undiff.[M]	1	1	3	1	.3078	.6493	.2551	.6751
stomach, glandular								
# Evaluated	60	55	56	60				
osteosarcoma[M]	0	0	1	0	.4653	.	.4561	.
testes								
# Evaluated	60	55	56	60				
adenoma, interst. cell[B]	2	0	0	0	1	1	1	1
thymus gland								
# Evaluated	49	46	45	55				
osteosarcoma[M]	0	1	0	0	.7059	.	.	.4444
thymoma[B]	1	0	0	0	1	1	1	1
thyroid gland								
# Evaluated	60	54	55	60				
adenoma, follicular cell[B]	0	1	0	0	.6900	.	.	.4259

Table A.3.6 Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

Organ/ Tumor	Incidence Cntr1	Low	Med	Hi	Significance Levels Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
cavity, abdominal								
# Evaluated	60	60	60	60				
adenocarcinoma[M]	1	0	1	0	.6728	1	.7385	1
cavity, thoracic								
# Evaluated	60	60	60	60				
adenocarcinoma[M]	0	0	1	0	.4298	.	.4923	.
harderian glands								
# Evaluated	60	59	60	60				
adenoma[B]	1	2	0	1	.5172	.5690	1	.4762
kidneys								
# Evaluated	60	60	60	60				
nephroblastoma[B]	1	0	0	0	1	1	1	1
liver								
# Evaluated	60	60	60	60				
Adenoma/Carcinoma; hepato.	2	1	5	2	.1123	.4196	.2005	.8751
adenoma, hepatocellular[B]	1	0	5	2	.0334	.2637	.0914	1
carcinoma, hepatocellular[M]	1	1	0	0	.9180	1	1	.7462
lung								
# Evaluated	60	60	60	60				
Adenoma/Carcinoma Bronch. Alv.	10	7	9	10	.0370	.0857	.6783	.8188
adenocarcinoma[M]	1	0	0	0	1	1	1	1
adenoma, bronch. alv.[B]	8	6	9	7	.0774	.1821	.5000	.7806
carcinoma, bronch. alv.[M]	2	1	0	4	.0336	.0932	1	.8634
sarcoma, undiff.[M]	0	0	1	0	.4348	.	.5000	.
lymph node, inguinal								
# Evaluated	60	60	60	60				
liposarcoma[M]	1	0	0	0	1	1	1	1

Table A.3.6 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

Organ/ Tumor	Incidence Cntr1	Low	Med	Hi	Significance Levels Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
lymph node, mediastinal								
# Evaluated	59	60	60	59				
adenocarcinoma[M]	1	0	0	0	1	1	1	1
lymph node, mesenteric								
# Evaluated	60	60	58	60				
adenocarcinoma[M]	0	0	1	0	.4298	.	.4923	.
mammary gland								
# Evaluated	60	60	60	60				
Adenocarcinoma/Sarcoma	2	0	1	0	.8147	1	.8692	1
adenocarcinoma[M]	2	0	0	0	1	1	1	1
sarcoma, undiff.[M]	0	0	1	0	.4298	.	.4923	.
multicentric neoplasm								
# Evaluated	60	60	60	60				
Hemagiosarcoma/Hemangioma	1	1	3	0	.5366	1	.3066	.7462
hemangioma[B]	0	1	1	0	.5131	.	.4923	.4923
hemangiosarcoma[M]	1	0	2	0	.5451	1	.5000	1
leukemia, granulocytic[M]	0	1	0	0	.7105	.	.	.4923
lymphoma[M]	5	3	6	2	.4900	.7788	.5212	.8507
sarcoma, histiocytic[M]	4	0	0	0	1	1	1	1
ovaries with oviducts								
# Evaluated	59	60	59	58				
adenocarcinoma[M]	1	0	0	0	1	1	1	1
cystadenoma[B]	1	2	0	0	.9019	1	1	.4883
leiomyosarcoma[M]	1	0	0	0	1	1	1	1
sex-cord/stromal tumor[B]	1	0	0	0	1	1	1	1
pancreas								
# Evaluated	60	60	60	60				
adenoma, islet cell[B]	0	0	1	0	.4348	.	.5000	.
pituitary gland								
# Evaluated	58	57	58	55				
Adenoma Pars dist./inter.	3	0	1	1	.5508	.7897	.9356	1
adenoma, pars distalis[B]	2	0	1	1	.4013	.6846	.8690	1
adenoma, pars intermedia[B]	1	0	0	0	1	1	1	1
skin								
# Evaluated	60	60	60	60				
carcinoma, basal cell[M]	0	0	1	0	.4298	.	.4923	.
skin, subcutis								
# Evaluated	60	60	60	60				
sarcoma, undiff.[M]	1	1	3	1	.2398	.5600	.2954	.7385
thyroid gland								
# Evaluated	59	59	60	59				
adenoma, follicular cell[B]	1	1	0	0	.9180	1	1	.7462
urinary bladder								
# Evaluated	60	59	60	60				
adenocarcinoma[M]	1	0	0	0	1	1	1	1

Table A.3.6 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

Organ/ Tumor	Incidence Cntr1	Low	Med	Hi	Significance Levels Trend	Hi vs Cntrl	Med vs Cntrl	Low vs Cntrl
uterus with cervix								
# Evaluated	60	60	57	60				
Leiomyoma [B]&[M]	4	0	0	0	1	1	1	1
Sarcoma/Polyp Stromal	6	3	4	0	.9488	1	.8079	.9108
adenocarcinoma[M]	2	0	1	0	.8097	1	.8630	1
granular cell tumor[B]	1	0	0	0	1	1	1	1
granular cell tumor[M]	1	0	0	0	1	1	1	1
leiomyoma[B]	2	0	0	0	1	1	1	1
leiomyosarcoma[M]	2	0	0	0	1	1	1	1
polyp, stromal[B]	4	3	3	0	.9140	1	.7465	.7606
sarcoma, stromal[M]	2	0	1	0	.8097	1	.8630	1
vagina								
# Evaluated	59	60	59	60				
granular cell tumor[M]	1	0	0	0	1	1	1	1

Appendix 4. FDA Parametric Bayesian Tumorigenicity Analysis

The frequentist approach to testing in the presence of multiplicities is to adjust the type I error rate (i.e., the probability of rejecting a true hypothesis of no differences) for the number of tumors. For example, the Haseman-Lin-Rahman rules for the Peto/poly-k tests described in Section 1.3.1.4. and also noted in Appendix 3 above, are designed to control the type I error for tests of trend and for pairwise tests each at about a 10% error rate. The Bayesian approach is less tied to Type I error, and assesses the probability of each of the multiple events on the basis of all information in the trial, including other events. Thus basically, the hierarchical nature of the Bayesian models is used to adjust for multiplicity. That is, each assessment of linear dose effect is adjusted for the other assessments.

An argument can be made that the fact that these assessments are conditional on observed data allows one to specify analyses conditional on some data based criteria. The criterion used here was that the proportion of animals observed with tumors was higher in the treatment groups than in the vehicle group, plus a further criterion that there were at least 8 animals with tumor in the rat studies and at least 6 in the mice studies. Incidence tables for each tumor obeying these criteria are given in Tables A.4.1 and A.4.3, below, for rats and mice, respectively.

For this analysis, we define a two stage mixture model on the treatment parameters in a simple logit model for tests of trend in carcinogenicity over dose for each relevant tumor. That is, define p_{ijk} as the probability of tumor i in subject j in treatment group k. That is, with i = 1 to n_t tumors and j = 1 to n_s animals, and dose d_k, leaving the experiment at time t_j and subject effect δ_j:

$$\text{logit}(p_{ijk}) = \alpha_i + \beta_i d_k + \gamma_i t_j / t_{max} + \delta_j, \; k=1,..,4, \; i=1,\ldots, n_t, \; j=1,\ldots,n_s.$$

with random subject effect $\delta_j \sim N(\mu_\delta, \sigma_\delta^2)$. We assign model priors:

$$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$$

$$\beta_i \sim N(\mu_\beta, \sigma_\beta^2).$$

and,

$$\gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2), \text{ all for } i = 1, \ldots, n_s.$$

with $\mu_\alpha = \mu_\beta = \mu_\gamma = \mu_\delta = 0$, $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\gamma^2 = \sigma_\delta^2 = 10000$.

Note this only involves a test of trend, not tests of differences between the treatment groups. An alternative would be reasonable to postulate a mixed prior on the primary parameter of interest, with a point mass at zero. With a continuous prior elsewhere, the posterior probability at the point could be interpreted as the posterior probability that the parameter is 0. However, for comparability with potential nonparametric Bayesian analysis, this was not done here. These analyses were implemented in WinBUGS 1.4 (Lunn *et al*, 2000).

Simple summaries of the posterior distributions of the effect of dose for each analyzed tumor are given in Tables A.4.2 and A.4.4, below. The tumor index matches the tumor number in Tables A.4.1 and A.4.3. The mean, median, and standard deviations in the tables below provide the summary estimates from the posterior probability distributions of the simple linear trend parameter for each specified organ tumor combination. The values under the headings

"HPD Low 2.5%" and "HPD High 97.5%" are the lower and upper limits of a 95% posterior probability interval, i.e. a so-called credibility interval. That is, given the data, the posterior probability that the parameter is in the interval is 0.95. If 0 is in the interval this could be considered as good evidence that the parameter is not so different from 0, i.e., if use this as a decision rule we would not reject the hypothesis that the parameter is close to 0. When 0 is not in the interval, this can be considered as strong evidence that the parameter is not 0 and in fact is quite different from 0. This is analogous to the usual frequentist approach to testing hypotheses, however, instead of assessing the distance of a test statistic from 0, it assess the distance of the actual parameter from 0. Summaries are presented for both dose effect, beta[tumor #], and the time effect, gamma[tumor #]. Only the dose effect is of immediate interest.

Table A.4.1 Trend and Incidence of Tumors in Rats Used in Bayesian Analysis

Tumor #	Organ	Tumor	Incidence Veh	Low	Med	High
Male Rats						
1	PANCREAS	Adenoma/Carcinoma; islet cell	3	5	4	2
2		Adenoma: islet cell	3	4	2	2
3	PITUITARY	Adenoma/Carcinoma; pars distalis	60	33	38	33
4		Adenoma: pars distalis	58	31	37	33
5	SUBCUTANEOUS TISSUE	Fibroma	2	3	3	0
Female Rats						
1	LIVER	Adenoma: hepatocellular	1	1	3	6
2	MAMMARY GLAND	Adenocarcinoma	33	17	25	17
3		Adenoma	21	23	24	16
4		Adenoma/Adenocarcinoma	44	32	42	26
5	PITUITARY	Adenoma/Carcinoma; pars distalis	103	57	56	52
6		Adenoma: pars distalis	89	49	49	48
7	THYROID	Adenoma/Carcinoma; Foll. cell	1	3	1	11
8		Adenoma: follicular cell	1	2	1	8

Table A.4.2 Summaries of Posterior Distributions in Rats

Parameter [tumor #]	Mean	Standard Dev	HPD Low 2.5%	Median	HPD High 97.5%
Male Rats					
beta[1]	0.1957	0.1954	-0.1979	0.2008	0.5594
beta[2]	0.1509	0.2051	-0.2724	0.1578	0.55
beta[3]	0.08987	0.167	-0.2454	0.09382	0.4022
beta[4]	0.1285	0.1657	-0.2193	0.134	0.4423
beta[5]	0.002906	0.2477	-0.5165	0.01561	0.459
gamma[1]	3.989	5.644	-4.832	2.533	15.32
gamma[2]	-1.772	4.878	-12.53	-2.586	7.204
gamma[3]	-0.4097	2.997	-5.763	-0.5071	6.563
gamma[4]	-1.951	3.195	-7.844	-1.755	3.315
gamma[5]	-0.7168	5.009	-8.884	-1.477	9.141

Table A.4.2 (cont.) Summaries of Posterior Distributions in Rats

Parameter [tumor #]	Mean	Standard Dev	HPD Low 2.5%	Median	HPD High 97.5%
Female Rats					
beta[1]	0.4141	0.1777	0.06548	0.4177	0.7628
beta[2]	-0.005345	0.1465	-0.3196	0.003173	0.2567
beta[3]	0.08699	0.1478	-0.2254	0.09455	0.3537
beta[4]	0.03355	0.1445	-0.2727	0.04002	0.2943
beta[5]	0.01575	0.1667	-0.3323	0.0223	0.3215
beta[6]	0.06437	0.1498	-0.2521	0.07254	0.3311
beta[7]	0.5011	0.1737	0.1425	0.5035	0.8265
beta[8]	0.4659	0.1775	0.1088	0.4699	0.8087
gamma[1]	5.651	4.284	-1.675	5.473	13.86
gamma[2]	10.99	3.194	5.478	10.64	18.45
gamma[3]	2.421	2.972	-2.647	1.985	8.274
gamma[4]	5.98	2.368	0.5674	5.918	10.47
gamma[5]	-7.519	4.27	-15.25	-8.389	1.609
gamma[6]	-4.767	3.157	-10.68	-5.081	1.585
gamma[7]	6.989	3.986	-0.4321	6.433	15.92
gamma[8]	3.808	3.392	-3.46	3.9	10.54

Note that 0 is in each of the five posterior credible intervals for the trend parameter in male rats. That suggests that there is no strong evidence of a dose related trend for any of the five tumors analyzed in males. However, 0 is outside the credible interval for trend in each of the tumors hepatocellular adenoma of the liver, follicular cell adenoma of the thyroid and pooled follicular cell adenoma and carcinoma of the thyroid. That is the probability of a non-zero trend in each of these tumors is above 0.975, and, in fact the posterior probability of a dose related increase in the chance of developing a tumor is 0.99 or above. In female rats, note that 0 is near the middle of the remaining intervals for a linear dose response, again suggesting no strong evidence of a dose effect.

Table A.4.3 Trend and Incidence of Tumors in Mice Used in Bayesian Analysis

Tumor #	Organ	Tumor	Incidence Veh	Low	Med	High
Male Mice						
1	liver	adenoma, hepatocellular[B]	2	2	5	1
2	skin, subcutis	sarcoma, undiff.[M]	1	1	3	1
Female Mice						
1	liver	Adenoma/Carcinoma; hepato.	2	1	5	2
2	liver	adenoma, hepatocellular[B]	1	0	5	2
3	skin, subcutis	sarcoma, undiff.[M]	1	1	3	1

Table A.4.4 Summaries of Posterior Distributions in Mice

Parameter [tumor #]	Mean	Standard Dev	HPD Low 2.5%	Median	HPD Med 97.5%
Male Mice					
beta[1]	-0.08296	0.4472	-0.961	-0.08773	0.7968
beta[2]	0.04729	0.4596	-0.8866	0.04229	0.9562
gamma[1]	8.411	4.517	0.08107	8.099	17.21
gamma[2]	0.4451	3.656	-6.361	0.4503	7.659
Female Mice					
beta[1]	-0.03905	0.2357	-0.4936	-0.04907	0.4331
beta[2]	0.1203	0.2454	-0.35	0.1162	0.5915
beta[3]	-0.1851	0.2283	-0.6397	-0.1845	0.2752
gamma[1]	10.44	4.744	2.06	9.934	19.67
gamma[2]	9.1	4.767	0.7266	8.766	18.2
gamma[3]	-1.519	3.677	-7.91	-1.68	6.42

In mice, 0 is also within each credible interval, suggesting no strong evidence of a dose related trend in any of these tumors (in fact the highest posterior probability of a positive trend is in benign hepatocellular adenoma of the liver in female mice at about 0.69).

Tables A.4.5 - A.4.8 show the tumor incidence of those neoplasms not meeting the entry criteria for this Bayesian analysis.

Table A.4.5 Incidence of Tumors in Male Rats Not Used in Bayesian Analysis

Organ	Tumor	Veh	Low	Med	High
ABDOMEN	Lipoma	0	0	1	0
	Malignant schwannoma	0	2	0	0
	Osteosarcoma	1	0	0	0
	Benign pheochromocytoma	4	2	1	0
	Malignant pheochromocytoma	3	1	0	0
	Pheochromocytoma [B]&[M]	7	3	1	0
BRAIN	Benign granular cell tumor	0	0	1	0
	Benign oligodendroglioma	1	0	0	0
	Malignant astrocytoma	3	1	1	2
	Malignant meningioma	1	0	0	0
	Papilloma: choroid plexus	0	1	0	0
CAVITY CRANIAL	Malignant schwannoma	0	0	0	1
CAVITY ORAL	Carcinoma: squamous cell	1	0	0	0
CECUM	Adenocarcinoma	0	0	1	0
EYE	Benign melanoma: uvea	1	0	0	0
HEMOLYM. TISSUE	Histiocytic sarcoma	1	2	1	1
	Leukemia: large granular lymphocyte	1	0	0	1
	Malignant lymphoma	0	0	1	2
	Mast cell tumor	0	1	0	0
JEJUNUM	Adenocarcinoma	1	0	0	0
	Leiomyosarcoma	1	0	0	0
KIDNEY	Adenoma: tubular cell	0	0	0	1
	Carcinoma: transitional cell	0	0	0	1

Table A.4.5 (cont.) Incidence of Tumors in Male Rats Not Used in Bayesian Analysis

Organ	Tumor	Veh	Low	Med	High
LACRIMAL GLAND	Fibroma	0	1	0	0
LIVER	Adenoma: hepatocellular	2	1	1	0
MAMMARY GLAND	Adenocarcinoma	0	2	0	0
	Adenoma	0	1	0	0
	Adenoma/Adenocarcinoma	0	3	0	0
	Fibroadenoma	4	1	3	1
PANCREAS	Carcinoma: acinar cell	1	0	0	0
	Carcinoma: islet cell	0	1	2	0
PARATHYROID GLAND	Adenoma	1	0	0	1
PITUITARY	Carcinoma: pars distalis	2	2	1	0
	Pituicytoma: pars nervosa	0	1	0	0
RECTUM	Adenoma	0	0	1	0
SKIN MISCELLANEOUS	Adenoma: basal cell	1	0	0	0
	Carcinoma: squamous cell	1	0	0	0
	Hemangioma	1	0	0	0
	Keratoacanthoma	0	0	1	0
	Papilloma/carcinoma; Sq. cell	3	1	3	0
	Papilloma: squamous cell	2	1	3	0
SPINAL CORD CERVICAL	Malignant astrocytoma	0	0	1	0
STOMACH	Leiomyosarcoma	1	1	0	0
SUBCUTANEOUS TISSUE	Fibrosarcoma	3	1	2	1
	Hemangiosarcoma	1	0	1	0
	Malignant fibrous histiocytoma	1	0	0	0
Systemic	Hemangioma/-Sarcoma	2	0	1	1
TAIL	Hemangioma	0	0	0	1
	Papilloma: squamous cell	2	1	0	1
TESTIS	Adenoma/Carcinoma; Intrst. cell	5	2	1	2
	Adenoma: interstitial cell	4	1	1	2
	Carcinoma: interstitial cell	1	1	0	0
THYROID	Adenoma/Carcinoma; C-cell	3	1	0	2
	Adenoma/Carcinoma; Foll. cell	2	0	2	3
	Adenoma: C-cell	3	1	0	2
	Adenoma: follicular cell	1	0	1	3
	Carcinoma: follicular cell	1	0	1	0
URINARY BLADDER	Carcinoma: transitional cell	0	0	0	1

Table A.4.6 Incidence of Tumors in Female Rats Not Used in Bayesian Analysis

Organ	Tumor	Veh	Low	Med	High
ABDOMEN	Paraganglioma (M)	1	0	0	0
ADRENAL	Adenoma/Carcinoma; Cortical	1	1	2	1
	Adenoma: cortical	1	1	0	1
	Benign pheochromocytoma	3	1	1	1
	Carcinoma: cortical	0	0	2	0
	Malignant pheochromocytoma	1	0	0	0
	Pheochromocytoma [B]&[M]	4	1	1	1
BRAIN	Malignant ependymoma	1	0	0	0
CAVITY ORAL	Carcinoma: squamous cell	0	1	0	0
CLITORAL GLAND	Adenoma	1	0	0	0
DUODENUM	Adenoma	0	1	0	0
HEMOLYM. TISSUE	Histiocytic sarcoma	0	1	0	0
	Malignant lymphoma	0	0	1	1
KIDNEY	Adenoma: tubular cell	0	0	0	2
	Carcinoma: tubular cell	0	0	1	0
MAMMARY GLAND	Fibroadenoma	39	10	6	1
MUSCLE SKELETAL MI	Hemangiosarcoma	0	1	0	0
OVARY	Benign thecoma	1	0	0	0
PANCREAS	Adenoma/Carcinoma; islet cell	6	2	1	3
	Adenoma: acinar cell	1	0	0	0
	Adenoma: islet cell	6	0	0	2
	Carcinoma: islet cell	0	2	1	1
PARATHYROID GLAND	Carcinoma	0	0	0	1
PITUITARY	Carcinoma: pars distalis	14	8	7	4
SKIN MISCELLANEOUS	Adenoma: basal cell	0	1	0	0
	Benign schwannoma	0	0	0	1
	Carcinoma: squamous cell	0	1	0	0
	Leiomyoma	0	0	0	1
	Papilloma/carcinoma; Sq. cell	1	1	1	0
	Papilloma: squamous cell	1	0	1	0
STOMACH	Leiomyoma	1	0	0	0
	Leiomyosarcoma	1	0	0	0
	Papilloma: squamous cell	1	0	0	0
SUBCUTANEOUS TISSUE	Fibroma	1	1	0	1
	Lipoma	1	0	0	0
	Malignant schwannoma	0	1	0	1
Systemic	Hemangioma/-Sarcoma	0	1	0	0
TAIL	Papilloma: squamous cell	1	0	0	0
THYROID	Adenoma/Carcinoma; C-cell	1	3	1	0
	Adenoma: C-cell	1	2	1	0
	Carcinoma: C-cell	0	1	0	0
THYROID	Carcinoma: follicular cell	0	1	0	3
TONGUE	Benign granular cell tumor	0	0	0	1
UTERUS	Benign granular cell tumor	2	0	0	0
	Polyp: endometrial stromal	10	1	0	0
	Sarcoma/Polyp; endo. stromal	12	1	0	3
	Sarcoma: endometrial stromal	2	0	0	3
VAGINA	Benign granular cell tumor	2	0	0	0

Table A.4.7 Incidence of Tumors in Male Mice Not Used in Bayesian Analysis

Organ	Tumor	Veh	Low	Med	High
adrenal glands	adenoma, subcapsular cell[B]	2	0	0	0
	pheochromocytoma[B]	0	1	0	0
cavity, thoracic	osteosarcoma[M]	0	1	0	0
epididymides	schwannoma[M]	0	0	1	0
harderian glands	adenoma[B]	4	1	2	2
kidneys	adenoma, tubular cell[B]	1	0	0	0
	nephroblastoma[B]	0	0	1	0
liver	Adenoma/Carcinoma; hepato.	5	2	6	2
	carcinoma, hepatocellular[M]	3	0	1	1
	osteosarcoma[M]	0	1	0	0
lung	Adenoma/Carc. Bronch. Alv	12	11	6	8
	adenoma, bronch. alv.[B]	12	10	5	7
	carcinoma, bronch. alv.[M]	0	1	1	1
	osteosarcoma[M]	0	1	0	0
	sarcoma, undiff.[M]	1	1	1	0
multicentric neoplasm	Hemagiosarcoma/Hemangioma	6	2	1	1
	hemangioma[B]	1	0	0	0
	hemangiosarcoma[M]	5	2	1	1
	lymphoma[M]	3	3	1	2
	sarcoma, histiocytic[M]	2	2	0	0
pancreas	adenoma, islet cell[B]	0	0	1	1
pharynx	papilloma, squamous cell[B]	0	0	0	1
pituitary gland	Adenoma Pars dist./inter.	0	0	0	1
	adenoma, pars distalis[B]	0	0	0	1
skin	carcinoma, squamous cell[M]	0	0	0	1
skin, subcutis	fibrous histiocytoma[M]	0	1	0	0
stomach, glandular	osteosarcoma[M]	0	0	1	0
testes	adenoma, interst. cell[B]	2	0	0	0
thymus gland	osteosarcoma[M]	0	1	0	0
	thymoma[B]	1	0	0	0
thyroid gland	adenoma, follicular cell[B]	0	1	0	0

Table A.4.8 Incidence of Tumors in Female Mice Not Used in Bayesian Analysis

Organ	Tumor	Veh	Low	Med	High
cavity, abdominal	adenocarcinoma[M]	1	0	1	0
cavity, thoracic	adenocarcinoma[M]	0	0	1	0
harderian glands	adenoma[B]	1	2	0	1
kidneys	nephroblastoma[B]	1	0	0	0
liver	carcinoma, hepatocellular[M]	1	1	0	0
lung	Adenoma/Carc. Bronch. Alv.	10	7	9	10
	adenocarcinoma[M]	1	0	0	0
	adenoma, bronch. alv.[B]	8	6	9	7
	carcinoma, bronch. alv.[M]	2	1	0	4
	sarcoma, undiff.[M]	0	0	1	0
lymph node, inguinal	liposarcoma[M]	1	0	0	0
lymph node, mediastinal	adenocarcinoma[M]	1	0	0	0
lymph node, mesenteric	adenocarcinoma[M]	0	0	1	0
mammary gland	Adenocarcinoma/Sarcoma	2	0	1	0
	adenocarcinoma[M]	2	0	0	0
	sarcoma, undiff.[M]	0	0	1	0
multicentric neoplasm	Hemagiosarcoma/Hemangioma	1	1	3	0
	hemangioma[B]	0	1	1	0
	hemangiosarcoma[M]	1	0	2	0
	leukemia, granulocytic[M]	0	1	0	0
	lymphoma[M]	5	3	6	2
	sarcoma, histiocytic[M]	4	0	0	0
ovaries with oviducts	adenocarcinoma[M]	1	0	0	0
	cystadenoma[B]	1	2	0	0
	leiomyosarcoma[M]	1	0	0	0
	sex-cord/stromal tumor[B]	1	0	0	0
pancreas	adenoma, islet cell[B]	0	0	1	0
pituitary gland	Adenoma Pars dist./inter.	3	0	1	1
	adenoma, pars distalis[B]	2	0	1	1
	adenoma, pars intermedia[B]	1	0	0	0
skin	carcinoma, basal cell[M]	0	0	1	0
thyroid gland	adenoma, follicular cell[B]	1	1	0	0
urinary bladder	adenocarcinoma[M]	1	0	0	0
uterus with cervix	Leiomyoma [B]&[M]	4	0	0	0
	Sarcoma/Polyp Stromal	6	3	4	0
	adenocarcinoma[M]	2	0	1	0
	granular cell tumor[B]	1	0	0	0
	granular cell tumor[M]	1	0	0	0
	leiomyoma[B]	2	0	0	0
	leiomyosarcoma[M]	2	0	0	0
	polyp, stromal[B]	4	3	3	0
	sarcoma, stromal[M]	2	0	1	0
vagina	granular cell tumor[M]	1	0	0	0

Appendix 5. References

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

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

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

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

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

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

Jara, A. (2007), "Applied Bayesian Non- and Semi-parametric Inference using DPpackage", *Rnews***, 7**, 3, 17-26.

Lin, K. K. and Ali, M.W. (2006), "Statistical Review and Evaluation of Animal Tumorigenicity Studies", *Statistics in the Pharmaceutical Industry,Third Edition,* edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.

Lin, K. K. and Rahman, M.A. (1998), "Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs", Journal of Biopharmaceutical Statistics. 8, 1, 1-15.

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

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

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). "Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments", *IARC Monographs on the Evaluation of Carcinogenic*

Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal, International Agency for Research Against Cancer, 311-426.

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

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

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

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

---

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

/s/

STEVEN F THOMSON
09/28/2011

KARL K LIN
09/28/2011
Concur with review

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 202107 **Applicant: Corcept Therapeutics** **Stamp Date:**

Drug Name: Corlux® (mifepristone) **NDA/BLA Type:** **Submission Date: 04/25/2011**

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

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			ISE not needed
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).		X		
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? __Y_____

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

Studies Submitted

C-1073-400: An Open-label Study of the Efficacy and Safety of CORLUX (mifepristone) in the Treatment of the Signs and Symptoms of Endogenous Cushing's Syndrome

Studies to be Reviewed

C-1073-400

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

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.				**FDA agreed**
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	**X**			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			**X**	
Appropriate references for novel statistical methodology (if present) are included.			**X**	

File name: 5_Statistics Filing Checklist for a New NDA_BLA110207

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

Safety data organized to permit analyses across clinical trials in the NDA/BLA.				
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			

Further Information Request

1. What is the parameter code for Diastolic Blood pressure (SAS variable name for the primary efficacy variable for the C-HT cohort)?

2. If you have provided subgroup analyses on Age, Gender, and Race, please let us know the location. If not, please provide those.

Japobrata Choudhury, Ph.D.	6-14-11
Reviewing Statistician	Date
Jon T. Sahlroot, Ph.D.	
Supervisor/Team Leader	Date

File name: 5_Statistics Filing Checklist for a New NDA_BLA110207

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

/s/

JAPOBRATA CHOUDHURY
06/15/2011

JON T SAHLROOT
06/16/2011