CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-995

STATISTICAL REVIEW(S)

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

NDA NO.:	21-995
SERIAL NO.:	Amendment
DATE RECEIVED BY THE CENTER:	September 21, 2006
DRUG NAME:	MK-0431
DOSAGE FORM:	Tablets
SPONSOR:	Merck Inc.
DOCUMENTS REVIEWED:	Electronic Copy Dated September 19, 2006
NAMES OF STATISTICAL REVIEWERS:	Yi Tsong, Ph.D. Meiyu Shen, Ph.D. Jinglin Zhong, Ph.D.
NAME OF CHEMISTRY REVIEWER:	Shah, Vibhakar J, Ph.D. Christine Moore, Ph.D.

Yi Tsong, Ph.D.
Deputy Director

Meiyu Shen, Ph.D., Mathematical Statistician

Jinglin Zhong, Ph.D., Mathematical Statistician

Concur:

Stella G. Machado
Director, DBVI

Distribution: NDA 21-995
OBVI/Y. Tsong, Ph.D.
OBVI/S.Machado Ph.D.
OB/Bob O'Neill, Ph.. D.
OB/ Patrician, Lillian, MS

ONDQA/DPE/ Shah, Vibhakar J, Ph.D.
ONDQA/DPE/ Christine Moore, Ph.D.

APPEARS THIS WAY
ON ORIGINAL

5 Page(s) Withheld

✓ § 552(b)(4) Trade Secret / Confidential

_____ § 552(b)(5) Deliberative Process

_____ § 552(b)(5) Draft Labeling

Withheld Track Number: Statistical- 1

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

/s/

MeiYu Shen
10/11/2006 02:42:02 PM
BIOMETRICS

Yi Tsong
10/12/2006 11:07:18 AM
BIOMETRICS

Jinglin Zhong
10/12/2006 02:54:03 PM
BIOMETRICS

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-995
Drug Name:	Januvia (Sitagliptin Phosphate) 25 qd, 50 qd and 100 mg qd
Indication(s):	An adjunct to diet and exercise to improve glycemic control in patients with type 2 diabetes mellitus as monotherapy and as combination therapy with metformin or a PPAR-y agonist (e.g. thiazolidinedione) when diet and exercise plus the single agent do not provide adequate glycemic control
Applicant:	Merck & Co., Inc.
Date(s):	Received 12/16/05; user fee (10 months) 10/16/06
Review Priority:	Standard
Biometrics Division:	Division of Biometrics 2
Statistical Reviewer:	Lee-Ping Pian, Ph.D.
Concurring Reviewers:	Todd Sahlroot, Ph.D. Biometrics Team Leader Thomas Permutt. Ph.D. Division Director
Medical Division:	Division of Metabolism and Endocrinology Products (DMEP)
Clinical Team:	Ilan Irony, M.D. Mary Parks, M.D. Division Director
Project Manager:	Lina Aljuburi, Pharm D.

Keywords: New Drug Application (NDA) review, clinical studies

Table of Contents

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

This NDA includes clinical trial data for MK-0431 (Januvia), a dipeptidyl peptidase 4 (DPP- 4) inhibitor for the treatment of Type 2 diabetes mellitus. The submission includes:

- Two 12-week Phase 2b studies in patients with Type 2 diabetes
- Four Phase 3 studies (18 to 24 week durations) in patients with Type 2 diabetes
- One 12-week Phase 3 study in Type 2 diabetics with chronic renal insufficiency. (This drug is substantially excreted by the kidney.)

See Table 1 for study summaries including doses studied. Of the four longer-term Phase 3 controlled efficacy and safety studies, two studies evaluated MK-0431 as monotherapy (one 18-week study (023) and one 24-week study (021)) and two studies evaluated MK-0431 as add-on therapy to metformin (020) or pioglitazone (019).

The Phase 2 studies (010 and 014) were well designed with sufficient power on the primary endpoint, HbA1c change from baseline, to assess dose response.

MK-0431 doses studied in Phase 2 ranged from 10 mg to 100 mg daily doses given once a day (qd) or as split doses (bid). 100 mg and 200 mg qd (the latter as monotherapy only) were studied in the Phase 3 trials in diabetic patients with normal renal function. The Phase 3 renal safety study assessed 25 mg and 50 mg qd. The sponsor has proposed a (single) MK-0431 dose of 100 mg qd in patients with Type 2 diabetes and normal renal function and reduced doses of 25 or 50 mg qd in patients with Type 2 diabetes complicated by renal insufficiency.

The phase 2 data showed that efficacy did not appear to be substantially affected if the drug was given as a single daily dose or as split (bid) doses. The 25, 50 and 100 mg doses were consistently superior to placebo in both Phase 2 studies. Maximal efficacy of the drug was achieved at 50 mg with no clear additional clinical benefit from 100 mg.

The proposed doses of 25 mg and 50 mg qd were shown to be effective in the renal-impaired population (study 028).

In the four Phase 3 studies, MK-0431 100 mg qd and 200 mg qd were statistically superior to placebo in HbA_{1c} change from baseline in patients with Type 2 diabetes. In the 24-week monotherapy study, the Least Squared Mean (LSM) differences from placebo (95% confidence intervals) in HbA_{1c} change from baseline were -0.79% (-0.96, -0.62) and -0.94% (-1.11, -0.77), respectively, for the 100 mg and 200 mg doses, respectively. In the 18-week monotherapy study, the differences were -0.60% (-0.82, -0.39) and -0.48% (-0.70, -0.26), respectively. Therefore the efficacy of the two doses overlapped in the two monotherapy studies. In the pioglitazone and the metformin add-on studies, the LSM differences between 100 mg and placebo in HbA_{1c} change from baseline were -0.70 (-0.85, -0.54), and -0.65 (-0.77, -0.53), respectively (Fig. 1). At the End-of-Phase 2 meeting, the Agency had questioned the sponsor's proposed MK-0431 daily dose of 200 mg in the Phase 3 studies. In retrospect, the MK-0431 daily 50 mg dose should have been included in the Phase 3 studies in patients with normal renal function.

In summary, 100 mg was shown to be efficacious as add-on therapy to metformin or pioglitazone. I recommend that, based on the Phase 2 efficacy results showing no clear lessening of clinical benefit for 50 mg compared to the proposed 100 mg dose, daily doses of 50 mg and 100 mg should both be made available to patients with Type 2 diabetes and normal renal function as monotherapy. For diabetic patients with renal insufficiency, the efficacy data suggest that doses of 25 mg and 50 mg are efficacious.

1.2 Overview of Clinical Program and Studies Reviewed

Table 1 Summary of Clinical Studies (Phase 2 and Phase 3)

Study ID Period (Phase 2 or Phase 3)	# of Center, Country	Total Sample Size	Patient population	Duration
021 7/04-7/05 (Phase 3)	111, 56 US[1] & PR[2] 16 Eu[3] 39 ROW[4]	100 mg 238 200 mg 250 placebo 253 ratio 1:1:1 Total 741	≥18 to ≤75 years of age (1) not on AHA[5] (off for ≥8 weeks) (2) on a single AHA (3) on a dual oral combination (≤50% maximal dose of both components) Rescue: metformin	Phase A: 24 wks Phase B:80 wks
023 10/04-8/05 (Phase 3)	114, 60 US 37 Eu 17 ROW	100 mg 205 200 mg 206 placebo 110 ratio 2:2:1 total 521	≥18 to ≤75 years of age (1) not on AHA (off for ≥8 weeks) (2) on a single AHA (3) on a dual oral combination (≤50% maximal dose of both components) Rescue: metformin	Phase A:18 wks Phase B:36 wks
019	71,	100 mg+pio(30/45 mg) 175	≥18 years of age	24 wks

Study ID Period (Phase 2 or Phase 3)	# of Center, Country	Total Sample Size	Patient population	Duration
7/04-9/05 (Phase 3)	28 US 11 Eu 11 ROW	Placebo+pio(30/45 mg) 178 Ratio 1:1 Total 353	(1) not on AHA (2) on a single AHA (PPAR or non-PPAR) (3) on dual oral combination (PPAR +AHA) Rescue: metformin	
020 7/04-7/05 (Phase 3)	100 46 US 25 Eu 29 ROW	100 mg + MF 464 Placebo+ MF 237 ratio 2:1 total 701	≥18 to ≤78 years of age (1) not on AHA (2) on a single AHA (metformin or other) (3) on dual oral combination (metformin+AHA) Rescue: pioglitazone	Phase A:24 wks Phase B:80 wks
028 12/04-10/05 (Phase 2)	57 US& PR 5 Eu 13 ROW	MK-0431 65 Stratum 1 (not on dialysis) 50 mg Stratum 2 (on dialysis) 25 mg Placebo 26 Ratio 2:1:MK-0431:placebo Total 91	≥18 years of age Type 2 diabetes and chronic renal insufficiency (1) not on oral AHA (off ≥8 weeks); or (2) on insulin monotherapy (3) on a single oral AHA; or on a low dose dual oral combination agent (≤50% maximal dose of both components)	Phase A:12 wks Phase B:42 wks
010 7/03-8/04 (Phase 2b dose ranging)	83 US 46 ROW	4 doses of MK-0431 (5, 12.5, 25, 50 mg bid), placebo, glipizide 5 mg (electively titrated to 10 mg bid) 743 randomized	21-75 years of age (1) not on AHA with HbA_{1c} ≥6.5 to <10% (2) on monotherapy HbA_{1c} ≥6 to ≤9%	12 weeks
014 9/03-7/04 (Phase 2b dose ranging)	59 US 65 ROW	4 doses of MK-0431(25, 50,100 mg qd or 50 bid), placebo 555 randomized	21-75 years of age (1) not on AHA with HbA_{1c} ≥6.5 to <10% (2) on monotherapy HbA_{1c} ≥6 to ≤9%	12 weeks

US[1]: United States, PR[2]: Purto Rico, Eu[3]: Europe, ROW[4]: Rest of the World, AHA[5]: Anti-Hyperglycemic Agent

The proposed indications for MK-0431 100 mg qd were: (1) as an adjunct to diet and exercise to improve glycemic control in patients with type 2 diabetes mellitus as monotherapy and (2) as combination therapy with metformin or a PPAR-γ agonist (e.g. thiazolidinedione) when diet and exercise plus the single agent do not provide adequate glycemic control.

2. INTRODUCTION

2.1 Overview

JANUVIA (sitagliptin phosphate; MK-0431) belongs to a new therapeutic class called dipeptidyl peptidase 4 (DPP-4) inhibitor which prolongs the half life of incretin hormones that regulate glucose homeostasis. The postulated effect of incretin hormones is to regulate glucose-dependent insulin secretion following their release into circulation from the gut in response to food intake.

For the four Phase 3 studies, two MK-0431 doses, 100 mg and 200 mg, were tested in the monotherapy trials and only one dose, 100 mg was tested in the add-on trials.

Patients in the monotherapy studies were ≥18 and ≤75 years of age with type 2 diabetes who either (1) not on AHA (Anti-Hyperglycemic Agent) (off for ≥8 weeks) or (2) on a single AHA; or (3) on low doses of dual oral combination agents (i.e., at ≤50% of maximal dose of both components).

For combination studies, patients were either: (1) not on AHA or (2) on a single AHA (background AHA or non background AHA) or (3) on dual oral combination treatment with background AHA and another AHA.

After screening (visit 1) patients either entered directly into the single-blind placebo run-in period (2 weeks, visit 2/3) or washout/dose titration/stabilization period (6-12 weeks, visit 2).

The randomization criterion for HbA_{1c} was ≥7% and ≤10% at, or within the 2 weeks prior to, the single-blind placebo run-in period, visit 3/week -2.

For patients not meeting specific glycemic goals during the treatment period, rescue medication was initiated. The rescue criteria were:

1. >270 mg/dL after Visit 4/Day 1 through Visit 6/Week 6,
2. >240 mg/dL after Visit 6/Week 6 through Visit 7/Week 12, and
3. >200 mg/dL after Visit 7/Week 12 up to (but not including) the final visit

The primary efficacy variable was HbA_{1c} change from baseline. The primary analysis population was the all-patients-treated (APT) population which included all randomized patients who have a baseline HbA1c and at least one post-randomization HbA1c. The imputation method to handle missing values was last observation carried forward (LOCF). For rescued patients the last HbA_{1c} prior to rescue was carried forward in the analysis. The secondary analysis was on the completers who had a baseline and an end of study HbA_{1c}.

The analysis used a covariance model which included treatment and AHA status as fixed effects and baseline HbA_{1c} as covariate. Although the prior AHA status was not a stratification factor, the sponsor included it in the ANCOVA model because it is a strong predictor of change in HbA_{1c}.

Figure 1 displays the placebo subtracted differences for MK-0431 by study. Figure 2 and figure 3 display HbA_{1c} changes from baseline and absolute HbA_{1c} over time, respectively for the APT population. The placebo groups had small increases in HbA_{1c} mean change from baseline (0.17, 0.16) for the monotherapy studies and small decreases (-0.18, -0.08) in the combination studies. 200 mg MK-0431 outperformed 100 mg MK-0431 in the 24-week monotherapy study (-0.94% vs. -0.79%) but not in the 18-week monotherapy study (-0.48% vs. -0.60%) in LSM difference from placebo in HbA_{1c} change. For the 100 mg MK-0431 proposed dose, the least squared mean differences from placebo were -0.79%, and -0.6% for the 24-week and 18-week monotherapy studies, respectively, and -0.7% and -0.65% for the pioglitazone and metformin add-on studies, respectively (Table 2). Similar to the APT analysis, the completers analyses were all statistically significant; however, the between group differences from the completers analysis were consistently less favorable than the APT analysis. The LSM differences between 100 mg and placebo were -0.65% and -0.48% for the 24-week, and 18-week monotherapy studies, respectively, and -0.63% and -0.55% for the pioglitazone add-on and metformin add-on studies, respectively, in the completers analysis (Table 3).

Table 2 HbA_{1c} change from baseline to study end – APT LOCF*

			Mean (SD)			LSM (SE)	LSM (CI)
	Treatment	N	Baseline	Endpoint	Change	Change	Difference from placebo
Monotherapy 24 wks	100 mg	229	8.01 (0.88)	7.39 (1.15)	-0.62 (1.02)	-0.61 (0.06)	**-0.79 (-0.96, -0.62)**
	200 mg	238	8.08 (0.94)	7.31 (1.14)	-0.78 (0.91)	-0.76 (0.06)	**-0.94 (-1.11, -0.77)**

			Mean (SD)			LSM (SE)	LSM (CI)
	Treatment	N	Baseline	Endpoint	Change	Change	Difference from placebo
	Placebo	244	8.03 (0.82)	8.2 (1.37)	+0.17 (1.06)	+0.18 (0.06)	
Monotherapy 18 wks	100 mg	193	8.04 (0.82)	7.58 (1.15)	-0.46 (0.85)	-0.48 (0.07)	**-0.60 (-0.82, -0.39)**
	200 mg	199	8.14 (0.91)	7.81 (1.31)	-0.34 (0.95)	-0.36 (0.06)	**-0.48 (-0.70, -0.26)**
	Placebo	103	8.05 (0.90)	8.21 (1.35)	+0.16 (0.93)	+0.12 (0.09)	
Add on Pioglitazone	100 mg	163	8.05 (0.81)	7.17 (0.91)	-0.88 (0.70)	-0.85 (0.07)	**-0.70 (-0.85, -0.54)**
	Placebo	174	8.00 (0.83)	7.82 (1.10)	-0.18 (0.79)	-0.15 (0.06)	
Add on Metformin	100 mg	453	7.96 (0.81)	7.26 (0.97)	-0.70 (0.72)	-0.67 (0.05)	**-0.65 (-0.77, -0.53)**
	Placebo	224	8.03 (0.82)	7.95 (1.10)	-0.08 (0.89)	-0.02 (0.06)	

*values prior to rescue medication were carried forward

Table 3 HbA_{1c} change from baseline to study end – Completers*

			Mean (SD)			LSM (SE)	LSM (CI)
	Treatment	N	Baseline	Endpoint	Change	Change	Difference from placebo
Monotherapy 24 wks	100 mg	189	7.92 (0.86)	7.13 (0.87)	-0.79 (0.89)	-0.76 (0.06)	-0.65 (-0.82, -0.49)
	200 mg	198	8.04 (0.87)	7.14 (0.95)	-0.90 (0.86)	-0.86 (0.06)	-0.75 (-0.91, -0.58)
	Placebo	176	7.88 (0.75)	7.76 (1.09)	-0.12 (0.93)	-0.11 (0.06)	
Monotherapy 18 wks	100 mg	168	7.96 (0.77)	7.4 (0.98)	-0.57 (0.77)	-0.59 (0.06)	-0.48 (-0.71, -0.26)
	200 mg	161	8.02 (0.86)	7.58 (1.19)	-0.43 (0.90)	-0.45 (0.07)	-0.34 (-0.57, -0.12)
	Placebo	74	7.9 (0.86)	7.83 (1.18)	-0.07 (0.87)	-0.10 (0.10)	
Add on Pioglitazone	100 mg	131	7.94 (0.73)	6.98 (0.75)	-0.95 (0.70)	-0.91 (0.07)	-0.63 (-0.79, -0.47)
	Placebo	136	7.86 (0.77)	7.57 (0.87)	-0.29 (0.73)	-0.28 (0.07)	
Add on Metformin	100 mg	399	7.91 (0.76)	7.12 (0.79)	-0.80 (0.66)	-0.81 (0.05)	-0.55 (-0.67,-0.44)
	Placebo	171	7.92 (0.76)	7.66(0.89)	-0.26 (0.84)	-0.26 (0.06)	

* patients with baseline and study end HbA_{1c}

Figure 1 LSM difference (C.I.) between MK-0431 and placebo by study - APT

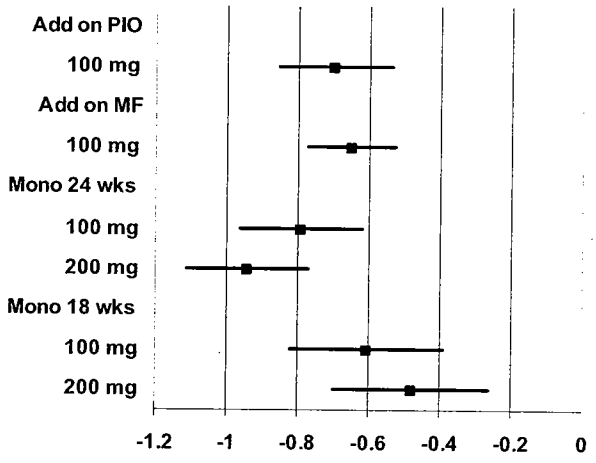
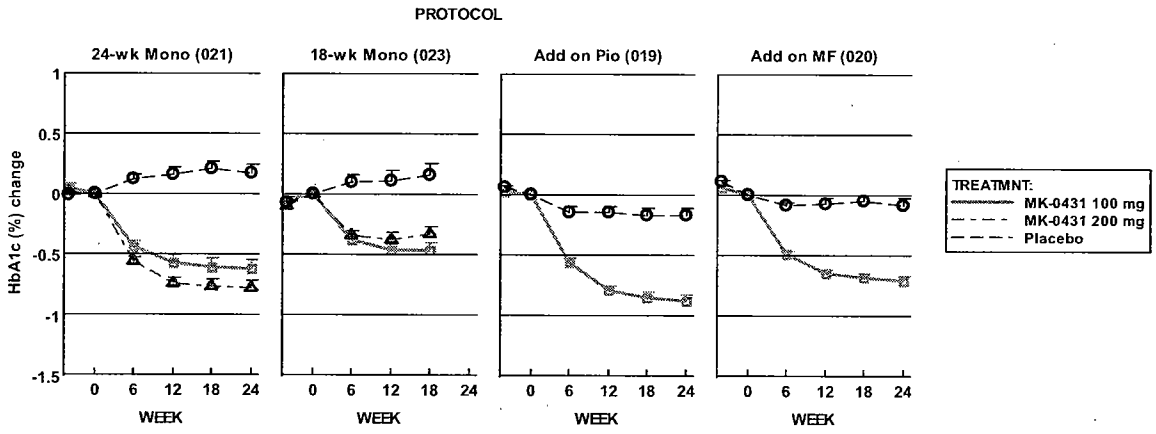
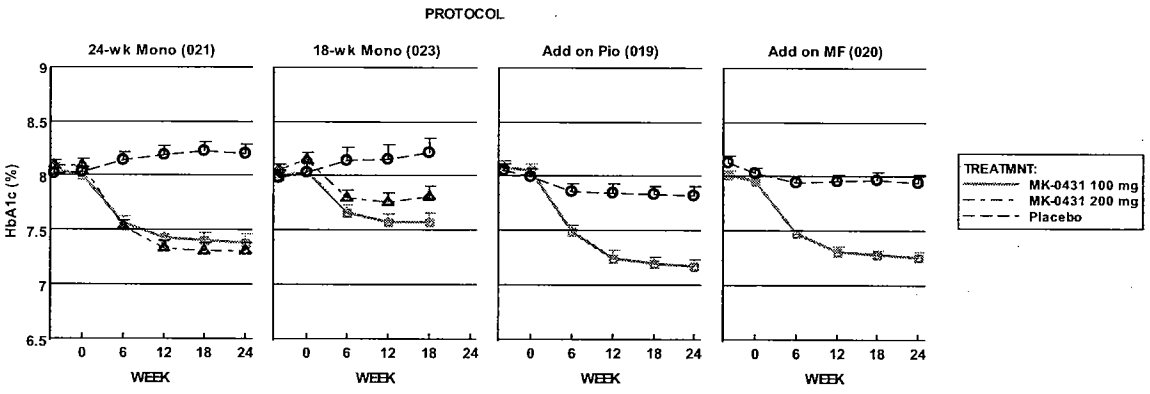


Figure 2 HbA_{1c} change from baseline over time – APT, LOCF



APPEARS THIS WAY
ON ORIGINAL

Figure 3 HbA_{1c} (%) over time – APT, LOCF



2.2 Data Sources

The eCTD is located at:\\CDSESUB1\EVSPROD\N021995. The location for datasets is at:\\cdsesub1\evsprod\N021995\0000\m5\datasets\. The data folders for the 4 Phase 3 trials were p019, p020v1, p021v1 and p023v1.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

Study Design and Endpoints

All 4 efficacy studies were multinational, double-blind, randomized and placebo-controlled to demonstrate efficacy of MK-0431 compared to placebo in glycemic control in patients with type 2 diabetes on diet and exercise. Patients not reaching specific glycemic goals during the placebo-controlled treatment period were to have rescue therapy initiated. The primary efficacy variable was HbA_{1c} change from baseline to endpoint.

The sample size was greater than planned since more patients were eligible for randomization after the long run-in period. The planned samples size was sufficient to detect with ≥98% power, a treatment mean difference of 0.5% (SD 1.0) in HbA_{1c} change from baseline with an α=0.05 and assuming a <10% dropout rate.

Table 4 displays the planned sample size and randomized patients for the 4 studies.

Table 4 Summary of sample size

Study	Design	Planned	Randomized	Ratio
021	24-wk monotherapy	600	741	1:1:1
023	18-wk monotherapy	500	521	2:2:1
019	Add-on pioglitazone	300	352	1:1
020	Add-on metformin	525	701	2:1

The efficacy analysis only used data before rescue medication. The sponsor did not perform sensitivity analysis using post rescue data.

3.1.1 Monotherapy – 021 (24-week)

The 2 monotherapy studies 021 and 023 included patients 18 to 75 years of age who were not currently on an AHA or on monotherapy (or low dose combination therapy at ≤50% of maximum dose of either agent). Patients with HbA_{1c} ≥7% and ≤10% were randomized in a ratio of 1:1:1 for P021 and 1:2:2 for P023 to placebo:100 mg: 200 mg MK-0431 qd The double-blind treatment period (phase A) was 24 weeks for Study 021 and 18 weeks for Study 023. The phase B study (not reported in this submission) was 80 weeks for Study P021 and 36 weeks for Study P023.

The design of the study included a screening diet/exercise run-in period of up to 15 weeks (including a 1-week screening period [Visits 1 to 2], up to a 12-week diet/exercise run-in period and antihyperglycemic agent (AHA) "wash-out" for patients on AHAs [Visits 2 to 3] and a 2-week single-blind placebo run-in period [Visits 3 to 4]) prior to randomization.

The monotherapy studies screened type 2 diabetes who were either (1) not on AHA (off for ≥8 weeks); or (2) on a single AHA; or (3) on low-dose dual oral combination therapy (i.e., at ≤50% of maximum labeled dose of both components). Eligible patients had HbA1c ≥7% and ≤10% at or within 2 weeks prior to Visit 3/Week -2. The exclusion criteria for fasting glucose was >260 mg/dL. Patients not on AHA with HbA1c ≥7% and ≤10% and met all other enrollment criteria were to directly enter the 2-week single-blind placebo run-in period at a combined Visit 2/3. Table 5 displays guidelines for run-in period management (Sponsor's table 9-1).

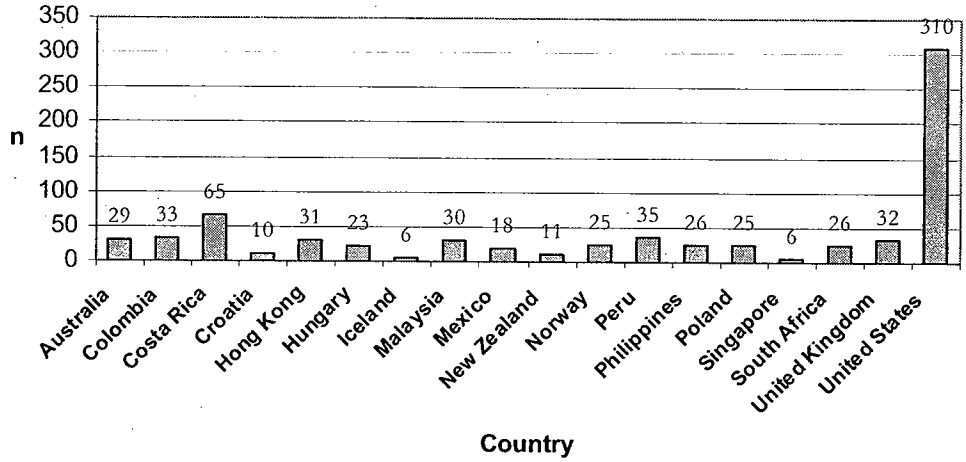
Table 5 Guidelines for Run-in Period Management

Only patients at Screening Visit/Visit 1 considered by the investigator as likely to meet Visit 3 HbA1c inclusion criteria (based upon evaluation of patient's current diet and exercise regimen, medication regimen, and the Visit 1 HbA1c level) may have been continued in study.		
Patient's Medication and Dose at Screening	Visit 1/Screening Visit HbA1c level	Visit 2 to Visit 3 Duration
Patient not on antihyperglycemic agent therapy (off for ≥8 weeks)	≥7% and ≤10%	Go directly to combined Visit 2/3
	>10%	Up to 6 weeks
On antihyperglycemic therapy (monotherapy or low dose combination)	≥7% and ≤10%	6 weeks Note: Patients on a TZD at Screening Visit/Visit 1 must have 8 weeks between Visit 2 and Visit 3
	≥6% and <7%	6 to 10 weeks Note: Patients on a TZD at Screening Visit/Visit 1 may have up to 12 weeks with a minimum of 8 weeks between Visit 2 and Visit 3

Patient Disposition, Demographic and Baseline Characteristics – 24-wk monotherapy

A total of 1807 patients were screened and 741 patients were randomized at 104 sites worldwide. This exceeded the 600 protocol planned sample size. The US randomized 42% of the patients, Costa Rica, 9% and <5% for the rest 16 countries. Of the 310 patients randomized in the U.S. sites, 66% were White, 19% were Hispanic and 12% were Black.

Figure 4 Number of patients randomized by country



Approximately 14% randomized patients withdrew from the study (Table 6). The most frequent reason for withdrawal was patient consent (5.1%).

Table 6 Patient disposition – Monotherapy Study 021

	MK-0431 100 mg n=238	MK-0431 200 n=250	Placebo n=253	Total n=741
Completed Phase A*	209 (87.8%)	214 (85.6%)	216 (85.4%)	639 (86.2%)
Discontinued	29 (12.1%)	36 (14.4%)	37 (14.6%)	102 (13.8%)
Clinical AE	5 (2.1%)	4 (1.6%)	4 (1.6%)	13 (1.8%)
Lab AE	0	0	1 (0.4%)	1 (0.1%)
Lack efficacy	3 (1.3%)	5 (2.0%)	9 (3.6%)	17 (2.3%)
Lost to follow-up	5 (2.1%)	4 (1.6%)	2 (0.8%)	11 (1.5%)
Other	2 (0.8%)	3 (1.2%)	5 (2.0%)	10 (1.3%)
Pat. Moved	3 (1.3%)	1 (0.4%)	1 (0.4%)	5 (0.7%)
Pat. Withdrew consent	10 (4.2%)	17 (6.8%)	11 (4.4%)	38 (5.1%)
Protocol dev	1 (0.4%)	2 (0.8%)	4 (1.6%)	7 (0.9%)

* Including patients rescued but not discontinued

Percentages of patients rescued were 9% (21/237), 5% (12/250), and 21% (52/253), for 100 mg, 200 mg, and placebo, respectively.

Table 7 (from sponsor's Table 10-3) displays the accounting of patients in the efficacy analysis populations for HbA_{1c}.

Table 7 Patients accounting in efficacy analysis population

Total Randomized	MK-0431 100 mg	MK-0431 200 mg	Placebo	Total
	238	250	253	741
Included in the APT† Analysis	229 (96.2)	238 (95.2)	244 (96.4)	711 (96.0)
Included in the Completers Analysis	189 (79.4)	198 (79.2)	176 (69.6)	563 (76.0)
Excluded from the APT† Analysis	9 (3.8)	12 (4.8)	9 (3.6)	30 (4.0)
No Baseline Data	2 (0.8)	0 (0.0)	0 (0.0)	2 (0.3)
No On-treatment Data	7 (2.9)	12 (4.8)	9 (3.6)	28 (3.8)

Total Randomized	MK-0431 100 mg	MK-0431 200 mg	Placebo	Total
	238	250	253	741
Excluded from the Completers Analysis‡	40 (16.8)	40 (16.0)	68 (26.9)	148 (20.0)
Rescued Prior to Week 24§ No Data at Week 24‖	17 (7.1)	10 (4.0)	45 (17.8)	72 (9.7)
	23 (9.7)	30 (12.0)	23 (9.1)	76 (10.3)

† APT: All-Patients-Treated.

‡ The completers population is a subset of the APT population including all patients with Week 24 data.

§ Efficacy data obtained on a patient after initiation of rescue therapy are treated as missing.

‖ For patients not on rescue medication.

Table 8 (from sponsor's Table 10-4) displays patient demographics.

The mean age of patients was 54.2 in years. More males were in the 100 mg group (57%) than in the 200 mg group (47%). Approximately 50% were Caucasian, 24% were Hispanic, 14% were Asian and 5% were Black. The mean baseline weight was 84.6 kg.

Table 8 Patient demographics

Age (years)

Treatment	N	Mean	Sd	Median	Range	
MK-0431 100 mg	238	53.4	9.5	55.0	24.0	to 75.0
MK-0431 200 mg	250	54.9	10.1	55.0	18.0	to 75.0
Placebo	253	54.3	10.1	54.0	23.0	to 75.0
All	741	54.2	9.9	54.0	18.0	to 75.0

Gender

	Male	Female	Total
Treatment	N (%)	N (%)	N
MK-0431 100 mg	136 (57.1)	102 (42.9)	238
MK-0431 200 mg	117 (46.8)	133 (53.2)	250
Placebo	130 (51.4)	123 (48.6)	253
All	383 (51.7)	358 (48.3)	741

Race

	White	Black	Hispanic	Asian	Other	Total
Treatment	N (%)	N (%)	N (%)	N (%)	N (%)	N
MK-0431 100 mg	122 (51.3)	10 (4.2)	58 (24.4)	32 (13.4)	16 (6.7)	238
MK-0431 200 mg	132 (52.8)	12 (4.8)	53 (21.2)	37 (14.8)	16 (6.4)	250
Placebo	127 (50.2)	16 (6.3)	64 (25.3)	34 (13.4)	12 (4.7)	253
All	381 (51.4)	38 (5.1)	175 (23.6)	103 (13.9)	44 (5.9)	741

Baseline Body Weight (kg)

Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	238	85.0	18.4	83.3	44.5	to 138.7
MK-0431 200 mg	250	83.7	19.2	83.8	45.0	to 145.4
Placebo	253	85.0	18.1	83.3	49.9	to 137.1
All	741	84.6	18.5	83.4	44.5	to 145.4

Baseline Body Mass Index (kg/m2)

Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	237	30.3	5.2	29.7	20.4	to 43.3
MK-0431 200 mg	250	30.3	5.4	29.8	19.1	to 43.0
Placebo	252	30.8	5.5	30.1	20.2	to 44.7
All	739	30.5	5.3	29.8	19.1	to 44.7

Table 9 (Table 10-5 from the sponsor) displays baseline efficacy characteristics. Baseline mean HbA_{1c} (SD) was 8.0% (0.9). The range was 6.3 to 10.9%. Approximately 50% of patients used AHA at baseline. Median duration of type 2 diabetes was 3 years (range 0 to 38 years).

Table 9 Baseline characteristics

Baseline HbA1c (%)

Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	236	8.0	0.9	7.8	6.6 to 10.9
MK-0431 200 mg	250	8.1	0.9	7.9	6.3 to 10.5
Placebo	253	8.0	0.8	7.9	6.6 to 10.7
All	739	8.0	0.9	7.9	6.3 to 10.9

Distribution of HbA1c at Baseline

Treatment	N	Number (%) of Patients with Baseline HbA1c		
		<8%	≥8 and <9%	≥9%
MK-0431 100 mg	236	135 (57.2)	62 (26.3)	39 (16.5)
MK-0431 200 mg	250	129 (51.6)	69 (27.6)	52 (20.8)
Placebo	253	132 (52.2)	85 (33.6)	36 (14.2)
All	739	396 (53.6)	216 (29.2)	127 (17.2)

Baseline FPG (mg/dL)

Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	238	170.7	43.0	165.0	94.0 to 427.0
MK-0431 200 mg	249	174.2	46.2	171.0	86.0 to 409.0
Placebo	253	176.1	41.8	172.0	73.0 to 283.0
All	740	173.7	43.7	168.5	73.0 to 427.0

Baseline Fasting Insulin (microIU/mL)

Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	233	14.5	13.1	11.8	0.1 to 138.9
MK-0431 200 mg	247	14.1	11.8	10.7	0.5 to 89.9
Placebo	252	14.9	10.8	12.2	1.5 to 76.0
All	732	14.5	11.9	11.5	0.1 to 138.9

Duration of Type 2 Diabetes Mellitus (years)

Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	238	4.3	4.9	3.0	0.0 to 38.0
MK-0431 200 mg	250	4.3	4.7	3.0	0.0 to 30.0
Placebo	252	4.6	4.7	3.0	0.0 to 35.0
All	740	4.4	4.8	3.0	0.0 to 38.0

Use of Anti-Hyperglycemic Medication at Screening

	Present	Absent	Total
Treatment	N (%)	N (%)	N
MK-0431 100 mg	114 (47.9)	124 (52.1)	238
MK-0431 200 mg	125 (50.0)	125 (50.0)	250
Placebo	124 (49.0)	129 (51.0)	253
All	363 (49.0)	378 (51.0)	741

Prevalence of Metabolic Syndrome†			
	Present	Absent	Total
Treatment	N (%)	N (%)	N
MK-0431 100 mg	139 (58.4)	99 (41.6)	238
MK-0431 200 mg	150 (60.0)	100 (40.0)	250
Placebo	169 (66.8)	84 (33.2)	253
All	458 (61.8)	283 (38.2)	741

Using the definition of the National Cholesterol Education Program

Results and Conclusions

The primary efficacy variable was HbA_{1c} change from baseline at Week 24. The secondary efficacy variables were change from baseline in Fasting Plasma Glucose (FPG) and change from baseline in 2-hour post-meal glucose at Week 24. For HbA_{1c} the hierarchical testing procedure compared the 100 mg group first then the 200 mg group to the placebo group. The testing procedure for the secondary efficacy variables proceeded in the order of FPG then 2-hour post-meal glucose conditioned on the prior test showed statistically significance at $\alpha=0.05$. Both treatment groups were statistically significantly better than placebo in HbA_{1c} change from baseline (Table 10, Fig. 5). The placebo subtracted least squared mean difference in HbA_{1c} change from baseline was -0.79 % for the 100 mg group and -0.94% for the 200 mg group. For completers, the differences from placebo were -0.65% and -0.75% for 100 mg and 200 mg. respectively (Table 11 and Fig. 6).

Table 10 ANCOVA results of HbA_{1c} (%) change from baseline at week 24 - 24-week Monotherapy

Treatment	N	Baseline Mean (SD)	Week 24 Mean (SD)	LSM change (SE)	Difference from placebo (CI)	p-value
100 mg	229	8.01 (0.88)	7.39 (1.15)	-0.61 (0.06)	-0.79 (-0.96,-0.62)	<0.001
200 mg	238	8.08 (0.94)	7.31 (1.14)	-0.76 (0.06)	-0.94 (-1.11,-0.77)	<0.001
Placebo	244	8.03 (0.82)	8.20 (1.37)	0.18 (0.06)		

ANCOVA model included factors treatment and prior AHA and baseline HbA_{1c} as covariate

Table 11 ANCOVA for HbA_{1c} (%) change from baseline at week 24 – Completers

		Mean (SD)		LSM (SE)		
Treatment	N	Baseline	Week 18	Change	Difference from placebo (CI)	p-value
100 mg	189	7.92 (0.86)	7.13 (0.87)	-0.76 (0.06)	-0.65 (-0.82, -0.49)	<0.001
200 mg	198	8.04 (0.87)	7.14 (0.95)	-0.86 (0.06)	-0.75 (-0.91, -0.58)	<0.001
Placebo	176	7.88 (0.75)	7.76 (1.09)	-0.11 (0.06)		

Figure 5 HbA_{1c} (%) change from baseline over time - APT

APPEARS THIS WAY
ON ORIGINAL

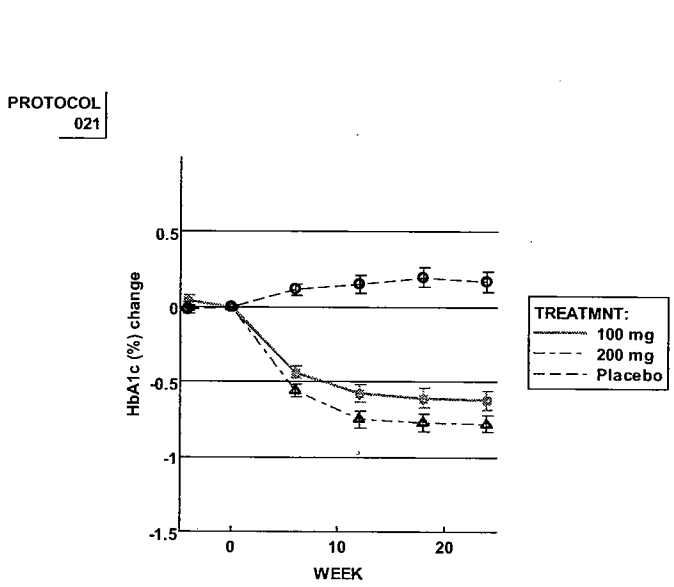
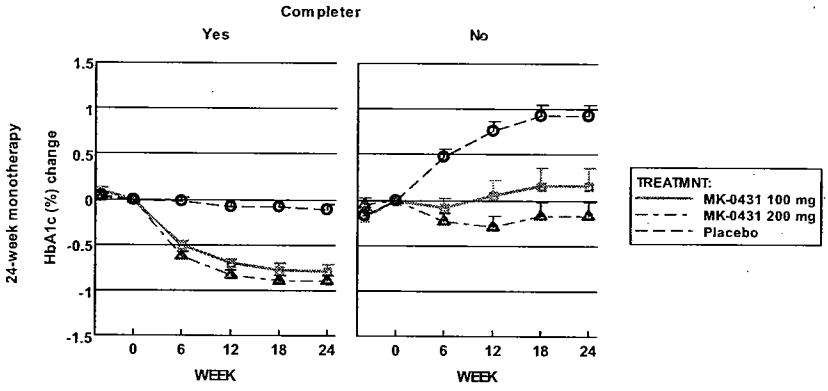


Figure 6 HbA_{1c} (%) change from baseline by completion status



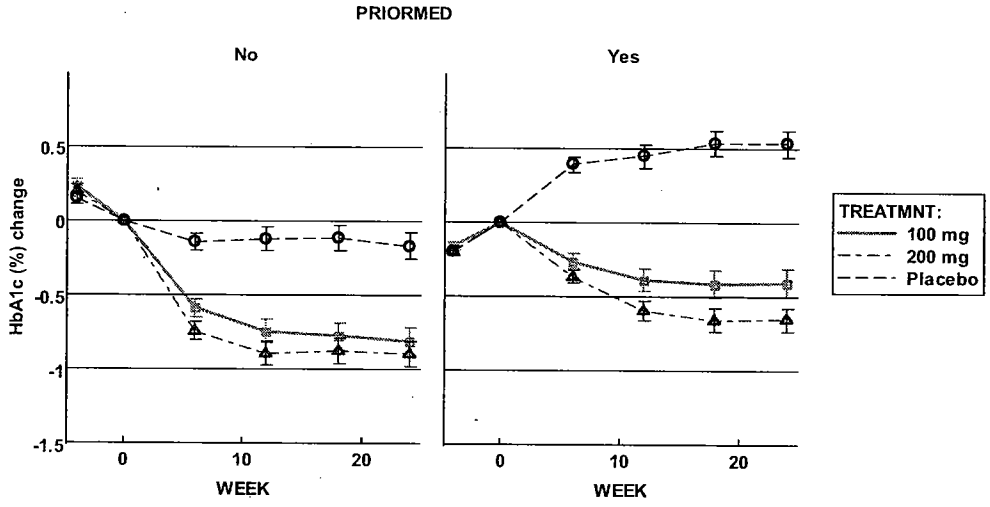
Treatment-by-prior AHA stratum interaction was significant (p=0.05). The LSM difference between MK-0431 group and placebo for each stratum is in Table 12. Figure 7 displays the HbA_{1c} change from baseline by stratum over time. The LSM increased over time only in placebo patients (0.55%) with their AHA washed out before randomization (Prior treated). Figure 8 displays that the interaction was quantitative but not qualitative in nature.

Table 12 HbA_{1c} change from baseline by prior medication stratum

TREATMNT	100 mg		200 mg		Placebo	
PRIORMED	No	Yes	No	Yes	No	Yes
n	(n=121)	(n=108)	(n=121)	(n=117)	(n=126)	(n=118)
LSM (StdErr)	-0.85 (0.09)	-0.38 (0.09)	-0.92 (0.09)	-0.61 (0.09)	-0.18 (0.08)	0.55 (0.09)
LSM diff (SE) from placebo	-0.66 (0.12)	-0.93 (0.13)	-0.74 (0.12)	-1.15 (0.12)		
LSM diff C.I.	(-0.90, -0.42)	(-1.18, -0.68)	(-0.97, -0.50)	(-1.40, -0.91)		

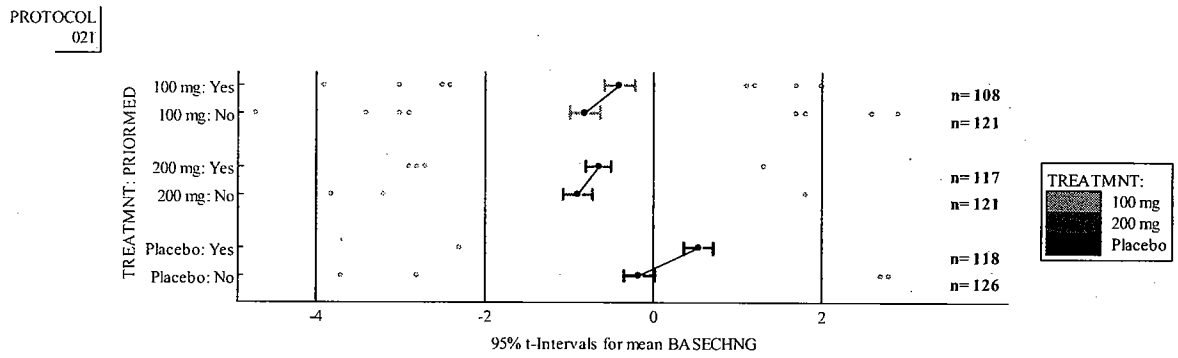
Figure 7 HbA_{1c} change from baseline over time stratified by prior AHA use

PROTOCOL
021



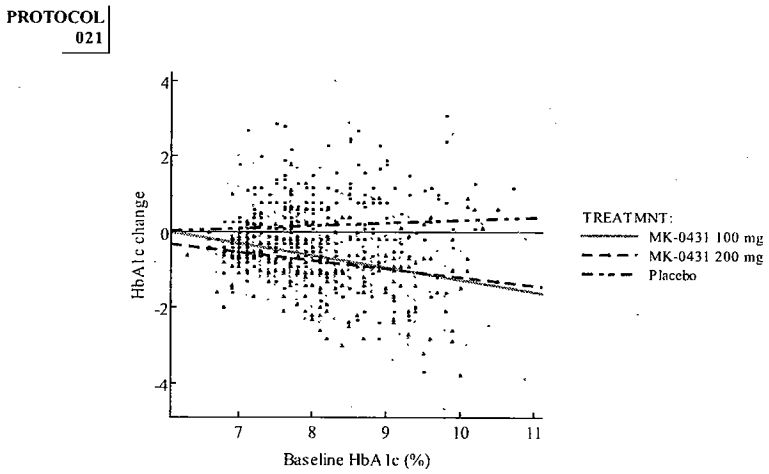
APPEARS THIS WAY
ON ORIGINAL

Figure 8 Mean (95% CI) in HbA_{1c} change from baseline at week 24 by treatment and prior AHA use



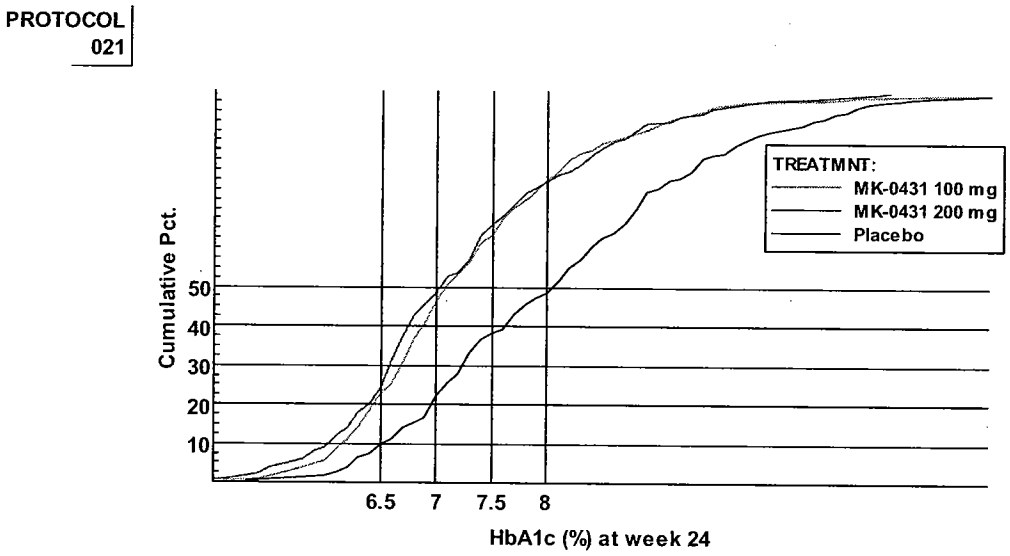
Treatment-by-baseline HbA_{1c} was significant ($p<0.001$). Figure 9 displays HbA_{1c} change from baseline at week 24 by baseline HbA_{1c}. Treatment difference between placebo and MK-0431 increased as baseline increased.

Figure 9 HbA_{1c} change from baseline at week 24 by baseline HbA1c



At week 24 (APT population), the percentages of patients with $HbA_{1c} \leq 7.0$ were 22%, 46% and 49%, respectively, for the placebo, MK-0431 100 mg and 200 mg treatment groups. For $HbA_{1c} \leq 6.5$ the percentages were 10%, 23% and 24%, respectively (Fig. 10).

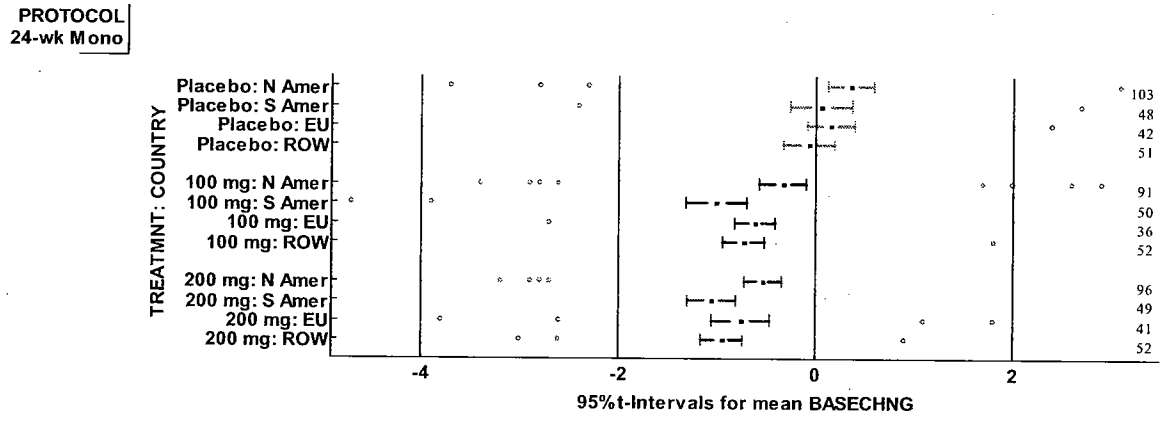
Figure 10 Cumulative distribution functions: HbA_{1c} (%) at week 24



Subgroup:

The treatment-by-region interaction was significant for North America and South America for the 100 mg group and the placebo group (p=0.1). The LSM difference between treatment in HbA_{1c} change from baseline was -1.09 (-1.52, -0.65) for S. America and -0.65 (-0.96, -0.34) for U.S (Fig. 11).

Figure 11 HbA_{1c} change from baseline at 24 weeks - APT



Race:

Treatment-by-race interaction was significant for White and Hispanic and Blacks (p<0.1) for the 100 mg and the placebo treatment groups (Figs. 12 & 13). The LSM difference between 100 mg and placebo in HbA_{1c} change from baseline was -1.4% (-2.2, -0.58) for Blacks, -1.0% (-1.37, -0.63) for Hispanics and -0.6% (-0.87, -0.36) for Whites. The estimates were consistent with the estimates from the subgroup for region.

Figure 12 HbA_{1c} change from baseline at 24 weeks - APT

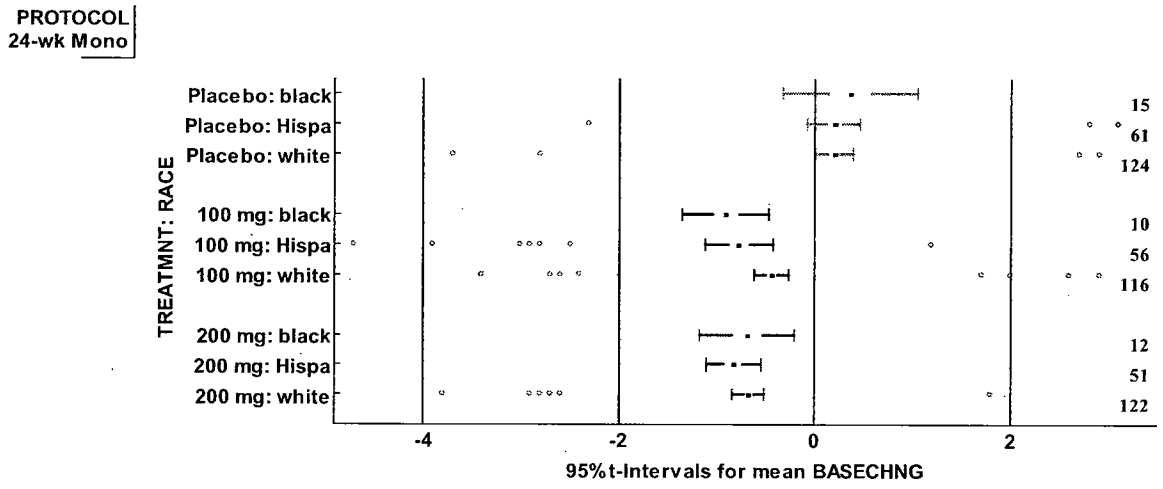
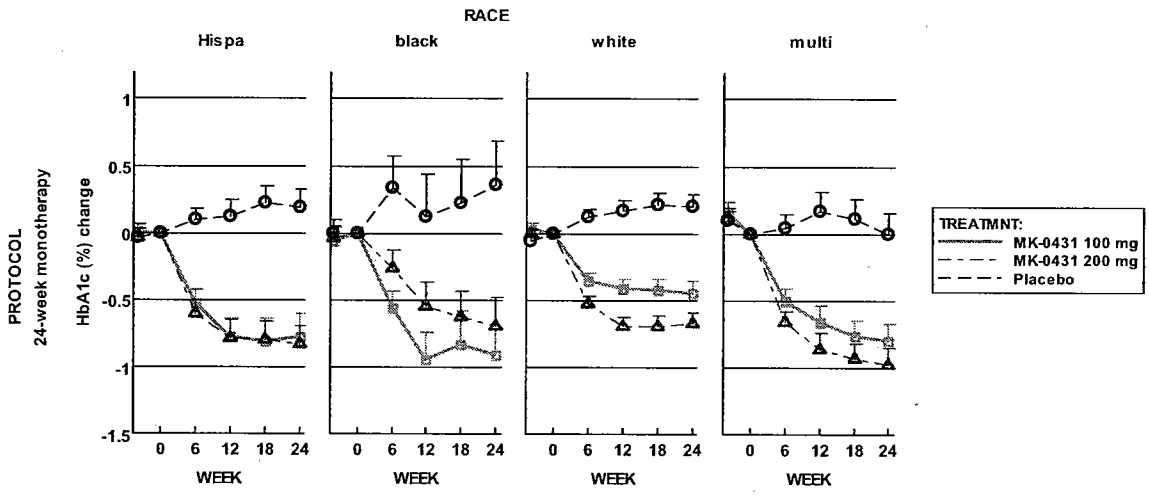


Figure 13 HbA_{1c} change over time by race



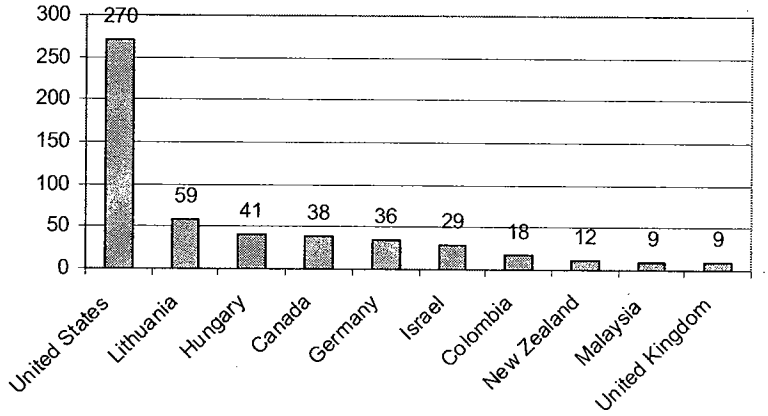
APPEARS THIS WAY
ON ORIGINAL

3.1.2 Monotherapy – 023 (18-week)

Patient Disposition, Demographic and Baseline Characteristics

A total of 1387 patients were screened and 521 patients were randomized at 97 centers. Figure 14 displays numbers of patients randomized by country. Of the 270 patients randomized at the U.S. sites, 52% of patients were White, 29% were Hispanic and 14% were Black.

Figure 14 Patients randomized by country



Approximately 89% of patients completed the study.

Table 13 Patient disposition – Monotherapy Study 023

	MK-0431 100 mg n=205	MK-0431 200 n=206	Placebo n=110	Total n=521
Completer	188 (91.7%)	184 (89.3%)	91 (82.7%)	463 (88.9%)
Discontinued	17 (8.3%)	22 (10.7%)	19 (17.3%)	58 (11.1%)
Clinical AE	1 (0.5%)	0	4 (3.6%)	5 (1.0%)
Lab AE	0	2 (1.0%)	0	2 (0.4%)
Lack efficacy	0	4 (1.9%)	6 (5.5%)	10 (1.9%)
Lost to follow-up	3 (1.5%)	3 (1.5%)	2 (1.8%)	8 (1.5%)
Other	1 (0.5%)	1 (0.5%)	0	2 (0.4%)
Pat. Moved	1 (0.5%)	1 (0.5%)	1 (0.9%)	3 (0.6%)
Pat. Withdrew consent	6 (2.9%)	7 (3.4%)	3 (2.7%)	16 (3.1%)
Protocol dev	5 (2.4%)	4 (1.9%)	3 (2.7%)	12 (2.3%)

* Including patients rescued but not discontinued

The numbers (%) of patients receiving rescue medication were 18 (8.8%), 24 (12%), and 19 (17%) for the 100 mg, 200 mg MK-0431 and placebo groups, respectively.

Table 14 (sponsor's Table 10-3) displays number of patients in the efficacy analysis populations for HbA_{1c}.

Table 14 Efficacy analysis patient population

Total Randomized	Number (%)			
	MK-0431 100 mg	MK-0431 200 mg	Placebo	Total
	205	206	110	521
Included in the APT† Analysis	193 (94.1)	199 (96.6)	103 (93.6)	495 (95.0)

Total Randomized	Number (%)			
	MK-0431 100 mg	MK-0431 200 mg	Placebo	Total
	205	206	110	521
Included in the Completers Analysis	168 (82.0)	161 (78.2)	74 (67.3)	403 (77.4)
Excluded from the APT†Analysis	12 (5.9)	7 (3.4)	7 (6.4)	26 (5.0)
No Baseline Data	3 (1.5)	1 (0.5)	1 (0.9)	5 (1.0)
No On-treatment Data	9 (4.4)	6 (2.9)	6 (5.5)	21 (4.0)
Excluded from the Completers Analysis‡	25 (12.2)	38 (18.4)	29 (26.4)	92 (17.7)
Rescued Prior to Week 18§	13 (6.3)	17 (8.3)	15 (13.6)	45 (8.6)
No Data at Week 18‖	12 (5.9)	21 (10.2)	14 (12.7)	47 (9.0)

† APT: All-Patients-Treated.

‡ The completers population is a subset of the APT population including all patients with Week 18 data.

§ Efficacy data obtained on a patient after initiation of rescue therapy are treated as missing.

‖ For patients not on rescue medication.

The mean age was 55 years. The placebo group had more males (63%) than females. Sixty-eight percent of patients were White. Of the 19% Hispanic patients, 15% were from the U.S. The mean baseline weight was 90 kg and the mean BMI was 32 kg/m^2 (Table 15).

Table 15 Baseline demographics

Age (years)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	205	54.5	10.0	54.0	29.0 to 75.0
MK-0431 200 mg	206	55.4	9.2	57.0	27.0 to 74.0
Placebo	110	55.5	10.1	55.0	27.0 to 76.0
All	521	55.1	9.7	56.0	27.0 to 76.0

Gender			
	Male	Female	Total
Treatment	N (%)	N (%)	N
MK-0431 100 mg	110 (53.7)	95 (46.3)	205
MK-0431 200 mg	104 (50.5)	102 (49.5)	206
Placebo	69 (62.7)	41 (37.3)	110
All	283 (54.3)	238 (45.7)	521

Race						
	White	Black	Hispanic	Asian	Other	Total
Treatment	N (%)	N (%)	N (%)	N (%)	N (%)	N
MK-0431 100 mg	142 (69.3)	16 (7.8)	37 (18.0)	8 (3.9)	2 (1.0)	205
MK-0431 200 mg	146 (70.9)	11 (5.3)	39 (18.9)	7 (3.4)	3 (1.5)	206
Placebo	68 (61.8)	12 (10.9)	22 (20.0)	5 (4.5)	3 (2.7)	110
All	356 (68.3)	39 (7.5)	98 (18.8)	20 (3.8)	8 (1.5)	521

Baseline Body Weight (kg)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	205	89.7	19.1	89.0	47.5 to 148.0
MK-0431 200 mg	206	89.6	19.4	87.9	52.0 to 144.7
Placebo	110	92.8	18.8	91.3	57.0 to 144.4
All	521	90.3	19.2	89.0	47.5 to 148.0

Baseline Body Mass Index (kg/m2)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	205	31.8	5.3	31.8	18.9 to 43.6
MK-0431 200 mg	205	32.0	5.3	31.7	19.8 to 42.9
Placebo	110	32.5	5.2	32.5	22.2 to 43.0
All	520	32.0	5.3	31.8	18.9 to 43.6

SD = Standard Deviation.

Mean baseline HbA_{1c} was approximately 8%. The mean duration of type 2 diabetes was 4.5 years (Table 16).

Table 16 Baseline characteristics – 18-week Monotherapy

Baseline HbA1c (%)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	202	8.0	0.8	7.9	6.2 to 10.3
MK-0431 200 mg	205	8.1	0.9	8.1	6.5 to 10.3
Placebo	109	8.0	0.9	7.9	6.6 to 10.5
All	516	8.1	0.9	7.9	6.2 to 10.5
Distribution of HbA1c at Baseline					
Treatment	N	Number (%) of Patients With Baseline HbA1c			
		<8%	≥8 and <9%	≥9%	
MK-0431 100 mg	202	103 (51.0)	70 (34.7)	29 (14.4)	
MK-0431 200 mg	205	99 (48.3)	62 (30.2)	44 (21.5)	
Placebo	109	63 (57.8)	26 (23.9)	20 (18.3)	
All	516	265 (51.4)	158 (30.6)	93 (18.0)	
Baseline FPG (mg/dL)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	205	180.2	43.3	174.0	98.0 to 335.0
MK-0431 200 mg	206	183.4	44.3	177.0	103.0 to 310.0
Placebo	110	183.7	48.5	172.5	92.0 to 306.0
All	521	182.2	44.8	174.0	92.0 to 335.0
Baseline Fasting Insulin (microIU/mL)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	199	14.7	8.9	13.1	1.1 to 56.4
MK-0431 200 mg	203	16.6	13.3	12.8	1.5 to 69.7
Placebo	109	17.5	16.8	13.4	3.1 to 140.7
All	511	16.0	12.7	13.1	1.1 to 140.7
Duration of Type 2 Diabetes Mellitus (years)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	204	4.5	4.3	3.0	0.0 to 23.0
MK-0431 200 mg	205	4.5	3.9	4.0	0.1 to 20.0
Placebo	110	4.7	5.0	3.5	0.1 to 30.0
All	519	4.5	4.3	3.0	0.0 to 30.0
Use of Anti-Hyperglycemic Medication at Screening					
	Present	Absent		Total	
Treatment	N (%)	N (%)		N	
MK-0431 100 mg	118 (57.6)	87 (42.4)		205	
MK-0431 200 mg	120 (58.3)	86 (41.7)		206	
Placebo	70 (63.6)	40 (36.4)		110	
All	308 (59.1)	213 (40.9)		521	
Prevalence of Metabolic Syndrome†					
	Present	Absent		Total	
Treatment	N (%)	N (%)		N	
MK-0431 100 mg	130 (63.4)	75 (36.6)		205	
MK-0431 200 mg	132 (64.1)	74 (35.9)		206	
Placebo	81 (73.6)	29 (26.4)		110	
All	343 (65.8)	178 (34.2)		521	

† Using the definition of the National Cholesterol Education Program [16.1.12.15]. SD = Standard Deviation.

Results and Conclusions

The primary efficacy variable was HbA_{1c} change from baseline and the secondary was Fasting Plasma Glucose (FPG). The LSM difference between MK-0431 and placebo in HbA_{1c} change from baseline to week 18 was statistically significant. The estimates were -0.60% (-0.82, -0.39) for the 100 mg group and -0.48% (-0.70, -0.26) for the 200 mg group (Table 17, Fig. 15). The estimates from the completer analysis were -0.48% and -0.34 for the 100 mg group and the 200 mg group, respectively (Table 18 and Fig. 16).

Table 17 ANCOVA for HbA_{1c} (%) change from baseline to week 18 – Monotherapy, Study 23

		Mean (SD)		LSM (SE)		
Treatment	N	Baseline	Week 18	Change	Difference from placebo (CI)	p-value
100 mg	193	8.04 (0.82)	7.58 (1.15)	-0.48 (0.07)	-0.60 (-0.82,-0.39)	<0.001
200 mg	199	8.14 (0.91)	7.81 (1.31)	-0.36 (0.06)	-0.48 (-0.70,-0.26)	<0.001
Placebo	103	8.05 (0.90)	8.21 (1.35)	0.12 (0.09)		

Table 18 ANCOVA for HbA_{1c} (%) change from baseline to week 18 – Completers

		Mean (SD)		LSM (SE)		
Treatment	N	Baseline	Week 18	Change	Difference from placebo (CI)	p-value
100 mg	168	7.96 (0.77)	7.4 (0.98)	-0.59 (0.06)	-0.48 (-0.71, -0.26)	<0.001
200 mg	161	8.02 (0.86)	7.58 (1.19)	-0.45 (0.07)	-0.34 (-0.57, -0.12)	
Placebo	74	7.9 (0.86)	7.83 (1.18)	-0.10 (0.10)		

Figure 15 Mean HbA_{1c} (%) over time - ITT

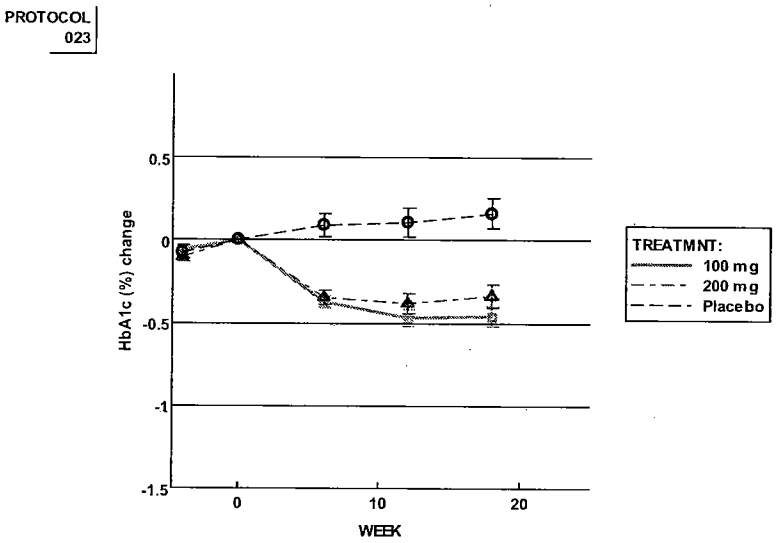
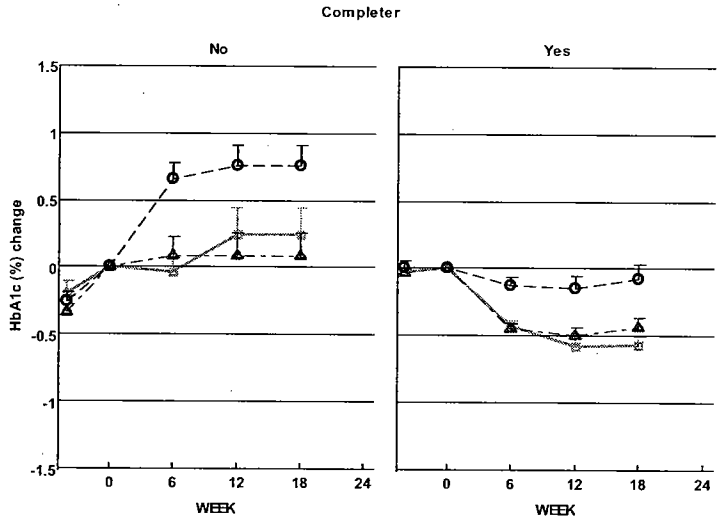


Figure 16 Mean HbA_{1c} (%) over time by completion status

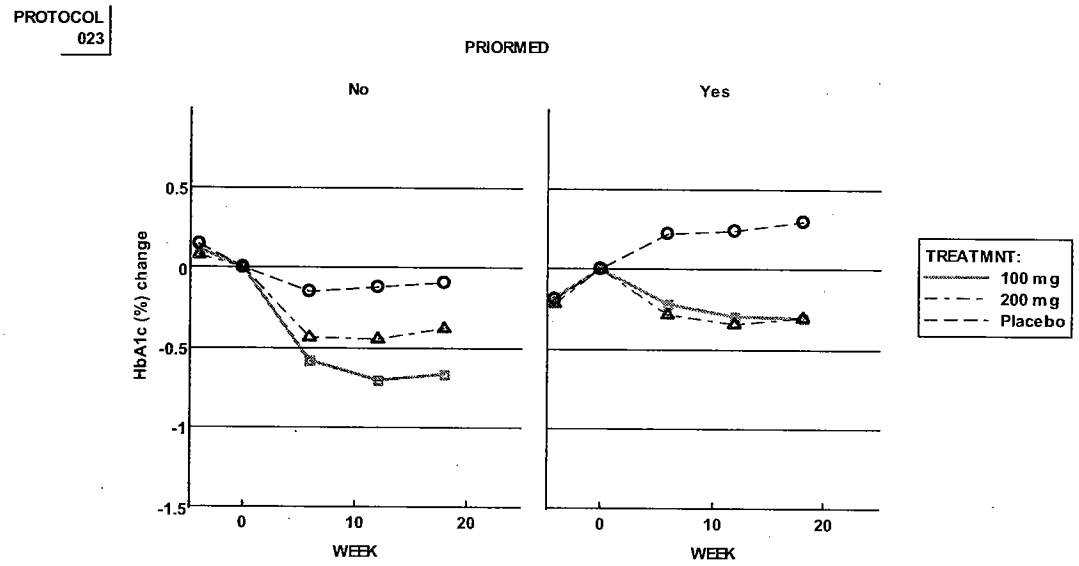
18-week monotherapy



APPEARS THIS WAY
ON ORIGINAL

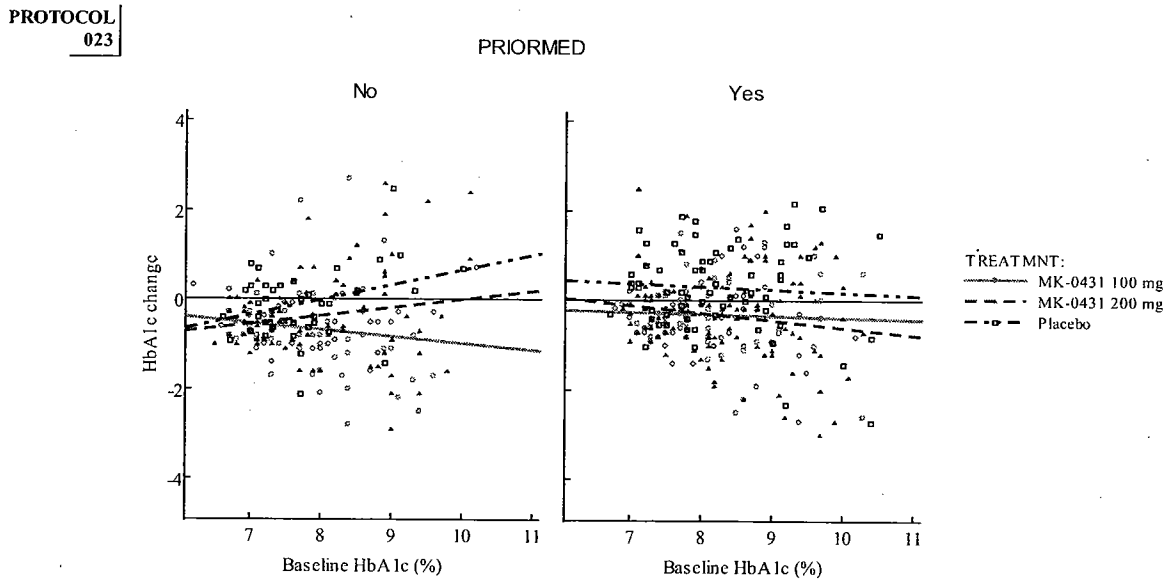
Figure 17 displays mean HbA_{1c} change from baseline by week. The baseline was not stabilized during the run in period for 'washed out', prior treated patients. As a result, the HbA_{1c} in the placebo patients continued to worsen. The difference between the 100 mg and the placebo was consistent in the 2 strata for prior AHA (no treatment-by-stratum interaction).

Figure 17 HbA_{1c} change from baseline over time stratified by prior AHA use



The treatment-by-baseline HbA_{1c} was significant ($p<0.1$) for 100 mg patients with no prior AHA use (Fig 18).

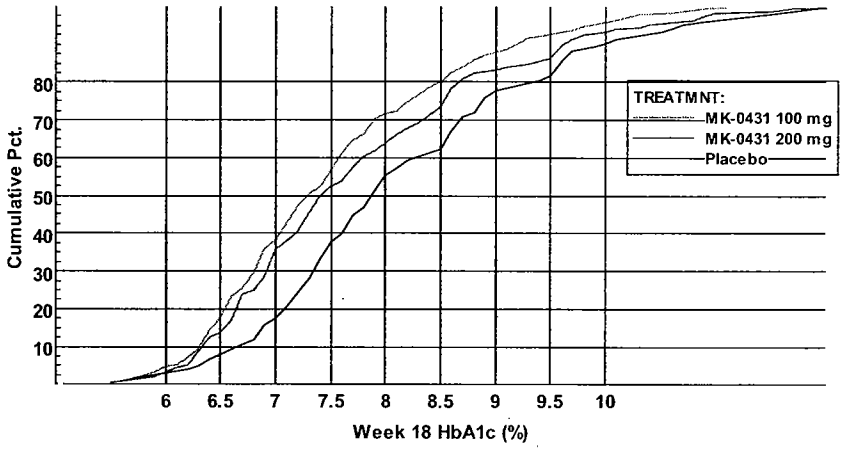
Figure 18 HbA_{1c} change from baseline to week 18 by baseline HbA_{1c} stratified by prior AHA use



The percentage of patients with $HbA_{1c} \leq 7$ were 18%, 38% and 36%, respectively for the placebo group, the 100 mg group and the 200 mg group, respectively. For $HbA_{1c} \leq 6.5\%$, the percentages were 8%, 18% and 14%, respectively (Fig. 19).

Figure 19 Cumulative distribution functions: HbA_{1c} at week 18

PROTOCOL 023



Add-on Study

Study Design and Endpoints

The 2 combination studies P019 and P020 included patients who were not currently on an AHA or on monotherapy or patients on dual oral combination therapies (the dual oral combination therapy had to include a PPARγ agonist in P019 and had to include metformin in P020). Patients with $HbA_{1c} \geq 7\%$ and $\leq 10\%$ were randomized in a ratio of 1:1 in study P019 and 1:2 in study P020 to placebo:MK-0431 100 mg qd Both studies were 24 weeks in duration (phase A). After Phase A period patients entered an 80-week double-blind treatment period (phase B). Placebo patients during phase B were switched to glipizide and MK-0431 100 mg patients continued on the same treatment.

Tables 19 and 20 display guidelines for screening run-in period for the two add-on studies.

Table 19 Guidelines for Run-in Period Management*

1. Only patients at Screening Visit/Visit 1 considered by the investigator as likely to meet Visit 3 HbA1c inclusion criteria (based upon evaluation of patient's current diet and exercise regimen, medication regimen, and the Visit 1 HbA1c level) may continue in the study. **2. At Visit 2 the patient's current antihyperglycemic agent(s) (if any) will be discontinued. At Visit 2 (or Visit 2b, 2-4 weeks after Visit 2), therapy with SPONSOR-supplied open-label pioglitazone will be initiated.** **3. For patients started on a dose of pioglitazone 30 mg, the dose may be increased to 45 mg within 4 weeks after starting pioglitazone 30 mg, but the dose-stable period duration must meet requirements indicated below.**		
Patient's Medication and Dose at Screening	Visit 1/Screening Visit HbA1c level	Pioglitazone Dose-Stable Period Prior to Visit 3
Patient not on anti-hyperglycemic agent therapy	>8%	14 weeks
Pioglitazone 30 to 45 mg (treated for ≥14 weeks)	≥7% and ≤10% >10%	Go directly to combined Visit 2/3 Up to 6 weeks
Antihyperglycemic agent monotherapy, including rosiglitazone at any dose or pioglitazone 15 mg (treated	≥7%	8 weeks

for ≥8 weeks†)		
PPARγ-based combination therapy (e.g., PPARγ-agent and metformin or PPARγ-agent and sulfonylurea; treated for ≥8 weeks)	≥7% and ≤10%	6 weeks
	≥6% and <7%	6 to 10 weeks

*sponsor Table 9-1, Reference P019

Table 20 Guidelines for Run-in Period Management*

1. Only patients at Screening Visit/Visit 1 considered by investigator as likely to meet Visit 3 HbA1c inclusion criterion (based upon evaluation of patient's current diet and exercise regimen, medication regimen, and the Visit 1 HbA1c level) may have been continued in study.
2. At Visit 2, therapy with SPONSOR-supplied open-label metformin was initiated, and the patient's current antihyperglycemic agent(s) (if any) was discontinued.
3. For patients started on metformin, the dose could be increased to ≥1500 mg/day and may have been increased up to a maximum dose of 2500 mg/day (or 3000 mg/day where the maximum dose of metformin per the local label is 3000 mg/day) within 6 weeks after Visit 2, but the dose-stable period duration must have met the requirements indicated below.

Patient's Medication and Dose at Screening	**Visit 1/Screening Visit HbA1c level**	Metformin Dose Stable Period Prior to Visit 3
Patient not on antihyperglycemic agent therapy	>8%	10 weeks
Metformin ≥1500 mg/day (treated for ≥10 weeks)	≥7% and ≤10%	Go directly to combined Visit 2/3
	>10%	Up to 6 weeks
Antihyperglycemic agent monotherapy (treated for ≥6 weeks, including metformin at <1500 mg/day)	≥7%	6 weeks Note: Patients discontinuing a TZD, could have a dose stable period of 8 weeks
Metformin-based oral combination therapy (treated for ≥6 weeks with metformin and sulfonylurea or metformin and (TZD)	≥7% and ≤10%	6 weeks Note: Patients discontinuing a TZD, could have a dose stable period of 8 weeks
	≥6% and <7%	6 to 10 weeks Note: Patients discontinuing a TZD, could have a dose stable period of at least 8 weeks

* sponsor Table 9-1, Reference P020V1

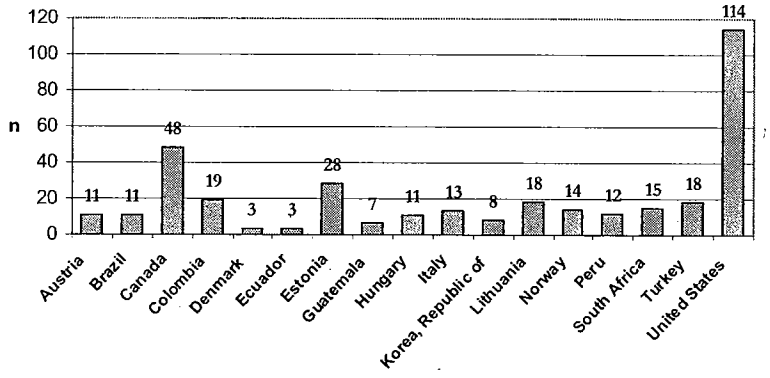
APPEARS THIS WAY
ON ORIGINAL

3.1.3 Add on to Pioglitazone - 019

Patient Disposition, Demographic and Baseline Characteristics

A total of 928 patients were screened and 353 patients randomized to 69 centers worldwide. Thirty-three percent of patients were from U.S. and 14% from Canada (Fig. 20). Of the 114 U.S. patients, 16% were Black, 11% were Hispanic, 4% were Asian and 68% were White.

Figure 20 Number of patients randomized by country



Approximately 87% (307) of patients completed the study. More patients in the MK-0431 group discontinued due to adverse events. (6.3% vs. 1.1%) (Table 21).

Table 21 Patient disposition – Add-on pioglitazone Study 019

	MK-0431 100mg+pioglitazone n=175	Plb + pioglitazone n=178	Total n=353
Completer	149 (85.1%)	158 (88.8%)	307 (87%)
Discontinued	26 (14.9%)	20 (11.2%)	46 (13%)
Clinical AE	11 (6.3%)	2 (1.1%)	13 (3.7%)
Lack efficacy	0 (0.0%)	2 (1.1%)	2 (0.6%)
Lost to follow-up	3 (1.7%)	1 (0.6%)	4 (1.1%)
Other	4 (2.3%)	5 (2.8%)	9 (2.5%)
Pat. Moved	1 (0.6%)	1 (0.6%)	2 (0.6%)
Pat. Withdrew consent	5 (2.9%)	6 (3.4%)	11 (3.1%)
Protocol dev	2 (1.1%)	3 (1.7%)	5 (1.4%)

* Including patients rescued but not discontinued

The number of patients receiving rescue medication was 25 (14%) for placebo and 12 (7%) for MK-0431 100 mg.

Table 22 (sponsor Table 10-3) displays the number of patients in the efficacy analysis populations for HbA_{1c}.

Table 22 Patient accounting – Add-on pioglitazone Study 019

	Number (%)		
Total Randomized	MK-0431 100 mg	Placebo	Total
	175	178	353

	Number (%)		
Total Randomized	MK-0431 100 mg	Placebo	Total
	175	178	353
Included in the APT† Analysis	163 (93.1)	174 (97.8)	337 (95.5)
Included in the Completers Analysis	131 (74.9)	136 (76.4)	267 (75.6)
Excluded from the APT† Analysis	12 (6.9)	4 (2.2)	16 (4.5)
No Baseline Data	1 (0.6)	0 (0.0)	1 (0.3)
No On-treatment Data	11 (6.3)	4 (2.2)	15 (4.2)
Excluded from the Completers Analysis‡	32 (18.3)	38 (21.3)	70 (19.8)
Rescued Prior to Week 24§	11 (6.3)	23 (12.9)	34 (9.6)
No Data at Week 24‖	21 (12.0)	15 (8.4)	36 (10.2)

† APT: All-Patients-Treated.
‡ The completers population is a subset of the APT population including all patients with Week 24 data.
§ Efficacy data obtained on a patient after initiation of rescue therapy are treated as missing.
‖ For patients not on rescue medication.

Patients mean age was 56 in years. There were more male patients (55%). Seventy-three percent patients were White, 12% were Hispanic and 7% were Black (Table 23).

Table 23 Patient baseline demographics – Add on Pioglitazone

Age(years)						
Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	175	55.6	10.4	55.0	31.0	to 80.0
Placebo	178	56.9	11.1	57.0	24.0	to 87.0
All	353	56.2	10.8	56.0	24.0	to 87.0

Gender			
Treatment	Male N (%)	Female N (%)	Total N
MK-0431 100 mg	93 (53.1)	82 (46.9)	175
Placebo	103 (57.9)	75 (42.1)	178
All	196 (55.5)	157 (44.5)	353

Race						
Treatment	White N (%)	Black N (%)	Hispanic N (%)	Asian N (%)	Other N (%)	Total N
MK-0431 100 mg	127 (72.6)	11 (6.3)	21 (12.0)	10 (5.7)	6 (3.4)	175
Placebo	129 (72.5)	12 (6.7)	22 (12.4)	5 (2.8)	10 (5.6)	178
All	256 (72.5)	23 (6.5)	43 (12.2)	15 (4.2)	16 (4.5)	353

Baseline Body Weight (kg)						
Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	175	90.9	17.0	89.0	50.5	to 133.5
Placebo	178	86.4	17.4	85.0	50.0	to 135.2
All	353	88.7	17.3	87.1	50.0	to 135.2

Baseline Body Mass Index (kg/m2)						
Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	175	32.0	5.2	31.7	20.1	to 44.2
Placebo	178	31.0	5.0	30.3	20.9	to 43.8
All	353	31.5	5.1	31.1	20.1	to 44.2

SD = Standard Deviation.

The mean HbA_{1c} at baseline was 8.0%. Mean duration of type 2 diabetes was 6.1 years (Table 24).

Table 24 Patient baseline characteristics – Add on Pioglitazone

Baseline HbA1c (%)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	174	8.1	0.8		6.4 to 10.1
Placebo	178	8.0	0.8	7.9 7.8	6.5 to 10.4

Baseline HbA1c (%)					
Treatment	N	Mean	SD	Median	Range
All	352	8.0	0.8	7.9	6.4 to 10.4
Distribution of HbA1c at Baseline					
Treatment	N	Number (%) of Patients with Baseline HbA1c			
		<8%		≥ 8 and <9%	≥ 9%
MK-0431 100 mg	174	88 (50.6)		56 (32.2)	30 (17.2)
Placebo	178	97 (54.5)		53 (29.8)	28 (15.7)
All	352	185 (52.6)		109 (31.0)	58 (16.5)
Baseline FPG (mg/dL)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	174	168.2	38.9	159.0	94.0 to 315.0
Placebo	178	165.3	39.8	155.5	97.0 to 305.0
All	352	166.8	39.3	158.0	94.0 to 315.0
Baseline Fasting Insulin (microIU/mL)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	169	9.7	5.7	8.3	2.0 to 35.8
Placebo	174	9.1	5.5	7.8	2.1 to 32.9
All	343	9.4	5.6	8.2	2.0 to 35.8
Duration of Type 2 Diabetes Mellitus (years)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	175	6.1	5.4	5.0	0.0 to 38.0
Placebo	178	6.1	5.7	5.0	0.0 to 32.0
All	353	6.1	5.6	5.0	0.0 to 38.0
Use of Anti-Hyperglycemic Medication at Screening					
Treatment	PPARγ-Based Combination Therapy		Monotherapy	Absence	Total
	N (%)		N (%)	N (%)	N
MK-0431 100 mg	52 (29.7)		109 (62.3)	14 (8.0)	175
Placebo	54 (30.5)†		103 (58.2)	20 (11.3)	177
All	106 (30.1)		212 (60.2)	34 (9.7)	352
Prevalence of Metabolic Syndrome‡					
	Present		Absent	Total	
Treatment	N (%)		N (%)	N	
MK-0431 100 mg	92 (52.6)		83 (47.4)	175	
Placebo	90 (50.6)		88 (49.4)	178	
All	182 (51.6)		171 (48.4)	353	
Prior PPARγ Status at Visit 1					
	On PPARγ		Not on PPARγ	Total	
Treatment	N (%)		N (%)	N	
MK-0431 100 mg	89 (50.9)		86 (49.1)	175	
Placebo	84 (47.2)		94 (52.8)	178	
All	173 (49.0)		180 (51.0)	353	

† AN 36331 used non-PPAR combination therapy at Screening.
‡ Using the definition of the National Cholesterol Education Program.
SD = Standard Deviation.

Efficacy HbA_{1c}.

At week 24, the between group difference in HbA_{1c} change from baseline was -0.7% for the APT population and -0.63% (Table 25 and Fig. 21) for the completers population (Table 26 & Fig. 22). Figure 23 displays the HbA_{1c} change over time by number of prior medications.

Table 25 ANCOVA results of HbA_{1c} (%) change from baseline -

Treatment	N	Baseline Mean (SD)	Week 24 Mean (SD)	LSM change (SE)	Difference from placebo (CI)	p-value
100 mg	163	8.05 (0.81)	7.17 (0.91)	-0.85 (0.07)	-0.70 (-0.85, -0.54)	<0.001
Placebo	174	8.00 (0.83)	8.20 (1.10)	-0.15 (0.06)		

ANCOVA model had treatment and prior AHA as factors and baseline HbA_{1c} as covariate

Table 26 ANCOVA for HbA_{1c} (%) change from baseline at week 24 – Completers

		Mean (SD)		LSM (SE)		
Treatment	N	Baseline	Week 18	Change	Difference from placebo (CI)	p-value
100 mg	131	7.94 (0.73)	6.98 (0.75)	-0.91 (0.07)	-0.63 (-0.79, -0.47)	<0.001
Placebo	136	7.86 (0.77)	7.57 (0.87)	-0.28 (0.07)		

Figure 21 HbA_{1c} over time – APT, LOCF

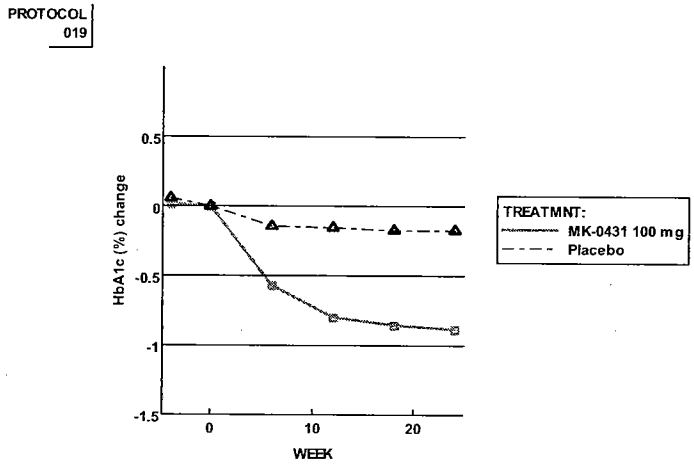


Figure 22 HbA_{1c} change over time by study completion

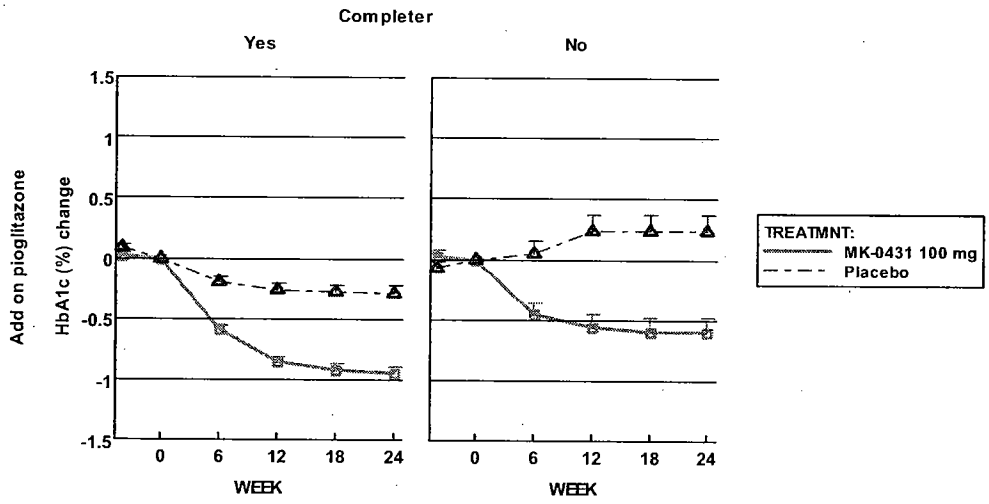
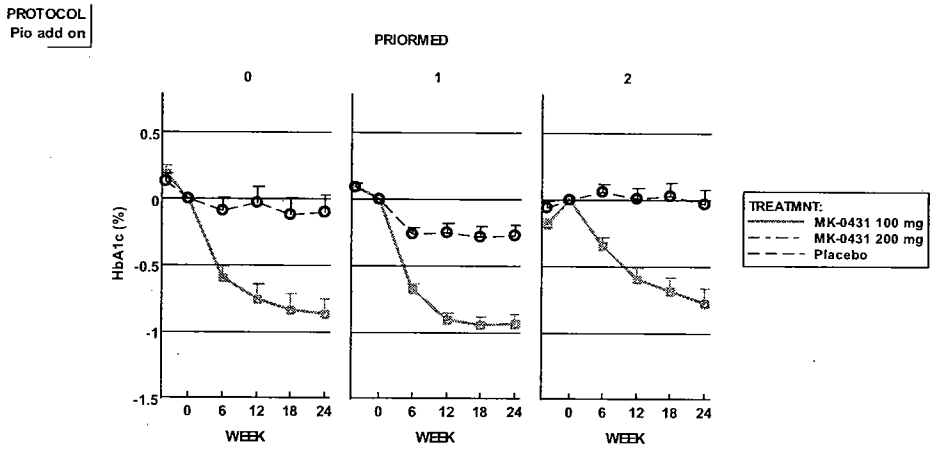
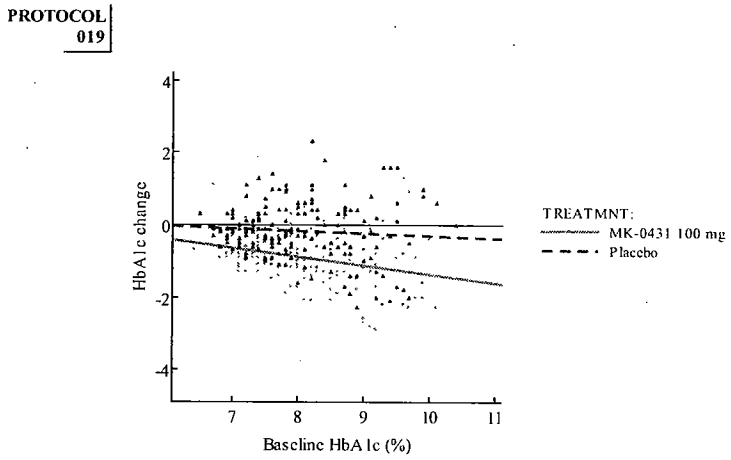


Figure 23 HbA_{1c} over time by number of prior medications



Treatment-by-baseline interaction was significant (p=0.05). Figure 24 displays the between group difference in HbA_{1c} change from baseline which increases as baseline HbA_{1c} increases.

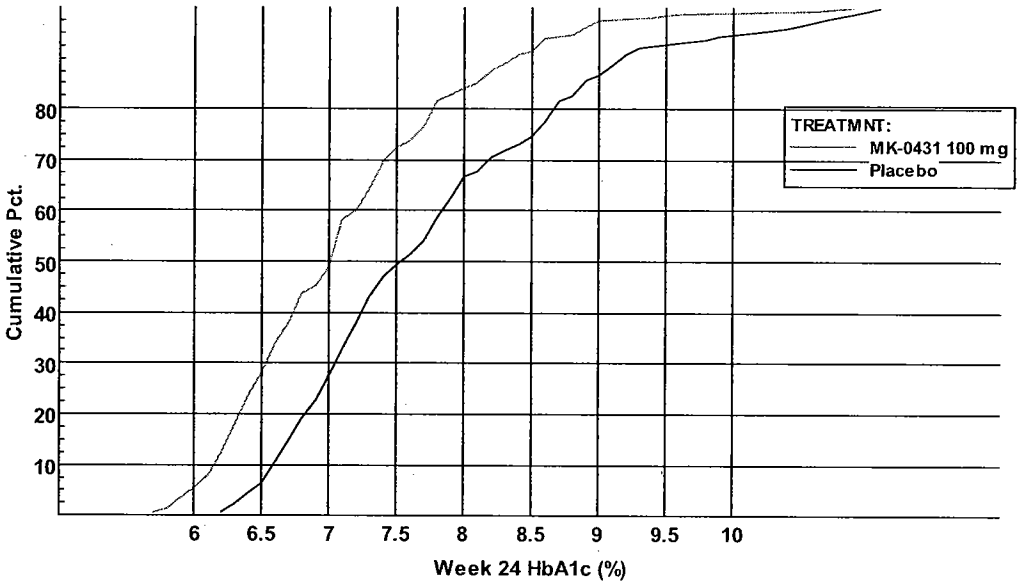
Figure 24 HbA_{1c} (%) change from baseline to week 24 by baseline HbA_{1c}



The percentages of patients with endpoint HbA_{1c} 7.0 were 28% (48/174) and 49% (80/163) for placebo and MK-0431, respectively. The percentages of patients with endpoint HbA_{1c} 6.5 were 6% (11/174) and 28% (46/163) in the placebo group and the MK-0431 100 mg group, respectively (Fig. 25).

Figure 25 Cumulative distribution functions: endpoint HbA_{1c} (%)

PROTOCOL 019



3.1.4 Add on to Metformin – 020

Patient Disposition, Demographic and Baseline Characteristics

A total of 1464 patients were screened and 701 were randomized at 99 sites worldwide. Approximately 41% of patients were at the U.S. sites (Fig. 26).

Figure 26 Number of patients randomized by country

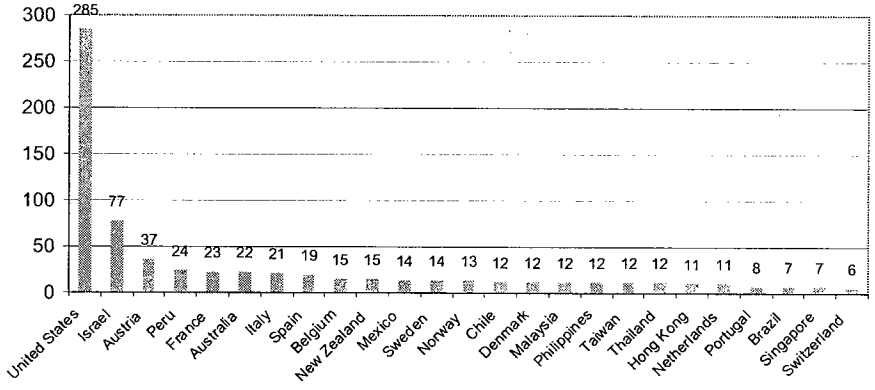


Table 27 Patient disposition – MF add on Study 020

	MK-0431 100mg+metformin n=464	Plb+metformin n=237	Total n=701

	MK-0431 100mg+metformin n=464	Plb+metformin n=237	Total n=701
Completed Phase A*	416 (89.7%)	192 (81.0%)	608 (86.7%)
Discontinued	48 (10.3%)	45 (19.0%)	93 (13.3%)
Clinical AE	11 (2.4%)	5 (2.1%)	16 (2.3%)
Lab AE	6 (1.3%)	4 (1.7%)	10 (1.4%)
Lack of efficacy	7 (1.5%)	13 (5.5%)	20 (2.9%)
Lost to follow-up	4 (0.9%)	5 (2.1%)	9 (1.3%)
Other	6 (1.3%)	4 (1.7%)	10 (1.4%)
Pat. Moved	2 (0.4%)	3 (1.3%)	5 (0.7%)
Pat. Withdrew consent	10 (2.2%)	10 (4.2%)	20 (2.9%)
Protocol dev	2 (0.4%)	1 (0.4%)	3 (0.4%)

* Including patients rescued but not discontinued

The percentages of rescued patients were 4.5% (21/464) and 13.5% (32/237) for 100 mg MK-0431 and placebo, respectively, in randomized groups.

Table 28 (sponsor Table 10-3) displays the accounting of patients in the efficacy analysis populations for HbA_{1c}. The lower percentage of placebo patients compared to 100 mg patients in the completers analysis was due to the higher placebo rescue rate (12% vs. 4%).

Table 28 Patient Accounting in the Analysis of HbA_{1c} at Week 24

	Number (%)		
	MK-0431 100 mg	Placebo	Total
Total Randomized	464	237	701
Included in the APT† Analysis	453 (97.6)	224 (94.5)	677 (96.6)
Included in the Completers Analysis	399 (86.0)	171 (72.2)	570 (81.3)
Excluded from the APT† Analysis	11 (2.4)	13 (5.5)	24 (3.4)
No Baseline Data	1 (0.2)	2 (0.8)	3 (0.4)
No On-treatment Data	10 (2.2)	11 (4.6)	21 (3.0)
Excluded from the Completers Analysis‡	54 (11.6)	53 (22.4)	107 (15.3)
Rescued Prior to Week 24§	18 (3.9)	28 (11.8)	46 (6.6)
No Data at Week 24‖	36 (7.8)	25 (10.5)	61 (8.7)

† APT: All-Patients-Treated.
‡ The completers population is a subset of the APT population including all patients with Week 24 data.
§ Efficacy data obtained on a patient after initiation of rescue therapy are treated as missing.
‖ For patients not on rescue medication.

Table 29 Patient baseline demographics – Add on Metformin

Age (years)						
Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	464	54.4	10.4	56.0	19.0	to 78.0
Placebo	237	54.7	9.7	56.0	26.0	to 76.0
All	701	54.5	10.2	56.0	19.0	to 78.0

Gender			
Treatment	Male N (%)	Female N (%)	Total N
MK-0431 100 mg	259 (55.8)	205 (44.2)	464
Placebo	141 (59.5)	96 (40.5)	237
All	400 (57.1)	301 (42.9)	701

Race						
Treatment	White N (%)	Black N (%)	Hispanic N (%)	Asian N (%)	Other N (%)	Total N
MK-0431 100 mg	293 (63.1)	31 (6.7)	72 (15.5)	49 (10.6)	19 (4.1)	464
Placebo	159 (67.1)	14 (5.9)	28 (11.8)	26 (11.0)	10 (4.2)	237
All	452 (64.5)	45 (6.4)	100 (14.3)	75 (10.7)	29 (4.1)	701
Baseline Body Weight (kg)						
Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	464	86.7	17.8	85.2	49.0	to 161.5
Placebo	237	89.6	17.5	88.5	53.0	to 146.7
All	701	87.7	17.7	86.1	49.0	to 161.5
Baseline Body Mass Index (kg/m2)						
Treatment	N	Mean	SD	Median	Range	
MK-0431 100 mg	464	30.9	5.3	30.1	19.6	to 43.9
Placebo	237	31.5	4.9	31.0	20.8	to 43.6
All	701	31.1	5.2	30.6	19.6	to 43.9
SD = Standard Deviation						

Table 30 Patient baseline characteristics – Add on Metformin

Baseline HbA1c (%)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	463	8.0	0.8	7.9	6.4 to 11.0
Placebo	235	8.0	0.8	7.9	6.4 to 10.3
All	698	8.0	0.8	7.9	6.4 to 11.0
Distribution of HbA1c at Baseline					
		Number (%) of Patients with Baseline HbA1c			
Treatment	N	<8%	≥8 and <9%	≥9%	
MK-0431 100 mg	463	253 (54.6)	146 (31.5)	64 (13.8)	
Placebo	235	128 (54.5)	71 (30.2)	36 (15.3)	
All	698	381 (54.6)	217 (31.1)	100 (14.3)	
Baseline FPG (mg/dL)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	464	170.2	40.9	163.0	86.0 to 312.0
Placebo	236	174.1	42.0	166.5	98.0 to 299.0
All	700	171.5	41.3	164.0	86.0 to 312.0
Baseline Fasting Insulin (microIU/mL)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	464	12.3	10.1	10.0	1.1 to 99.1
Placebo	236	12.4	7.6	11.0	1.0 to 53.9
All	700	12.3	9.4	10.3	1.0 to 99.1
Duration of Type 2 Diabetes Mellitus (years)					
Treatment	N	Mean	SD	Median	Range
MK-0431 100 mg	462	6.0	5.0	5.0	0.1 to 33.0
Placebo	237	6.6	5.5	5.0	0.1 to 34.0
All	699	6.2	5.2	5.0	0.1 to 34.0
Use of Anti-Hyperglycemic Medication at Screening					
Treatment	Combination Therapy N (%)	Monotherapy N (%)	Absence N (%)	Total N	
MK-0431 100 mg	160 (34.5)	277 (59.7)	27 (5.8)	464	
Placebo	69 (29.1)	154 (65.0)	14 (5.9)	237	
All	229 (32.7)	431 (61.5)	41 (5.8)	701	
Prevalence of Metabolic Syndrome†					

	Present	Absent	Total
Treatment	N (%)	N (%)	N
MK-0431 100 mg	282 (60.8)	182 (39.2)	464
Placebo	149 (62.9)	88 (37.1)	237
All	431 (61.5)	270 (38.5)	701

† Using the definition of the National Cholesterol Education Program
SD = Standard Deviation

The between group LSM difference in HbA_{1c} change from baseline was statistically significant for both the APT and the completers analyses. The estimates were -0.65% for the APT analysis and -0.55% for the completers analysis (Tables 31 & 32). Figures 27 and 28 display the HbA_{1c} change from baseline over time. Figure 29 displays the HbA_{1c} change by number of prior medication. The previous treated patients in the placebo group were not stabilized during the run-in/ 'wash out' period. The placebo HbA_{1c} continued to increase (priormed=2, Fig. 29).

Table 31 ANCOVA for HbA_{1c} (%) change from baseline – Add on Metformin, Study 20

		Mean (SD)		LSM (SE)		
Treatment	N	Baseline	Week 18	Change	Difference from placebo (CI)	p-value
100 mg	453	7.96 (0.81)	7.26 (0.97)	-0.67 (0.05)	-0.65 (-0.77,-0.53)	<0.001
Placebo	224	8.03 (0.82)	7.95(1.10)	-0.02 (0.06)		

Table 32 ANCOVA for HbA_{1c} (%) change from baseline to Week 24 – Completers

		Mean (SD)		LSM (SE)		
Treatment	N	Baseline	Week 18	Change	Difference from placebo (CI)	p-value
100 mg	399	7.91 (0.76)	7.12 (0.79)	-0.81 (0.05)	-0.55 (-0.67,-0.44)	<0.001
Placebo	171	7.92 (0.76)	7.66(0.89)	-0.26 (0.06)		

Figure 27 HbA_{1c} change from baseline over time - APT

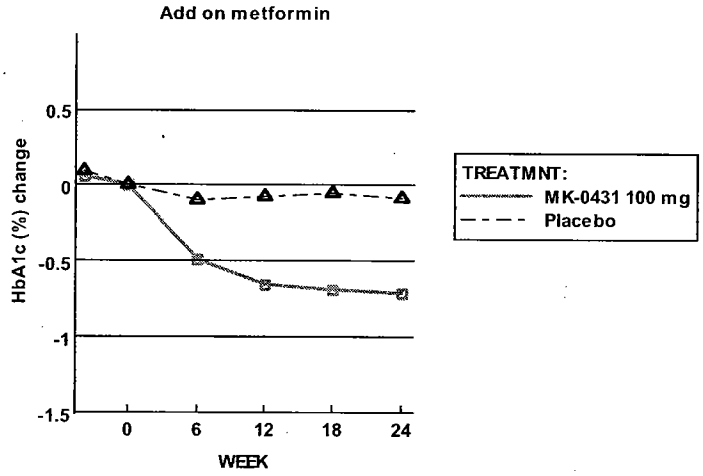


Figure 28 HbA_{1c} change from baseline over time by study completion

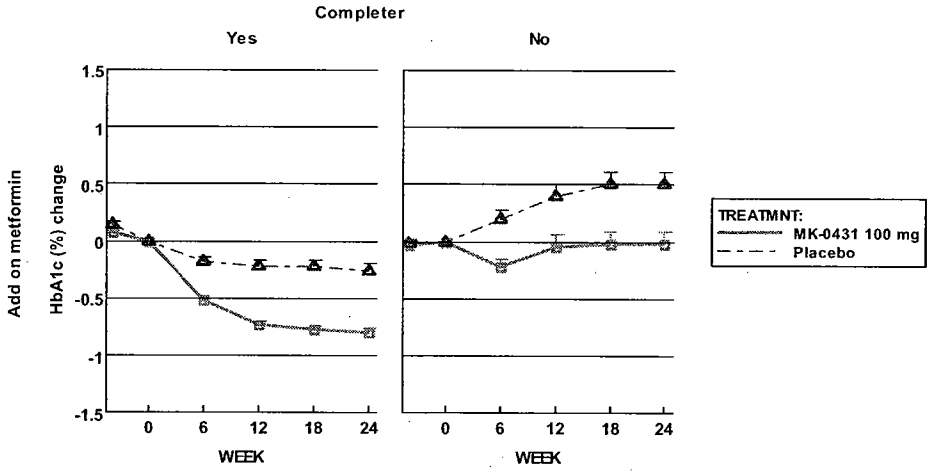
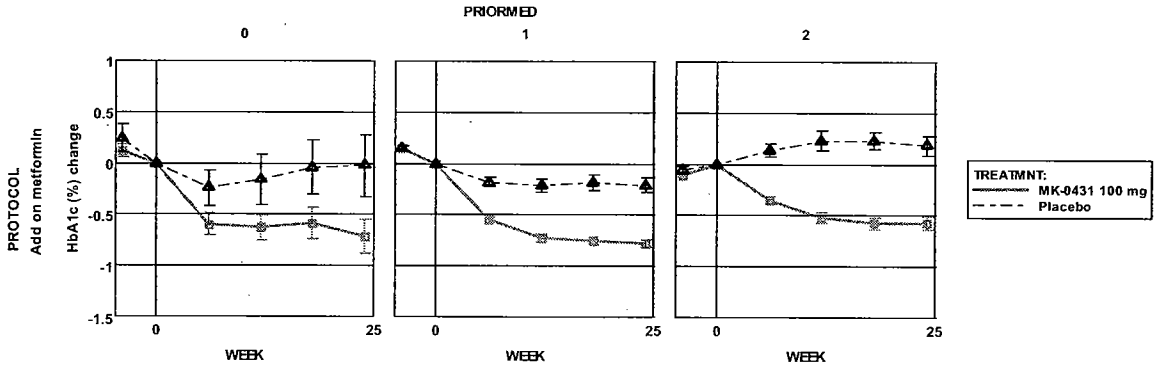


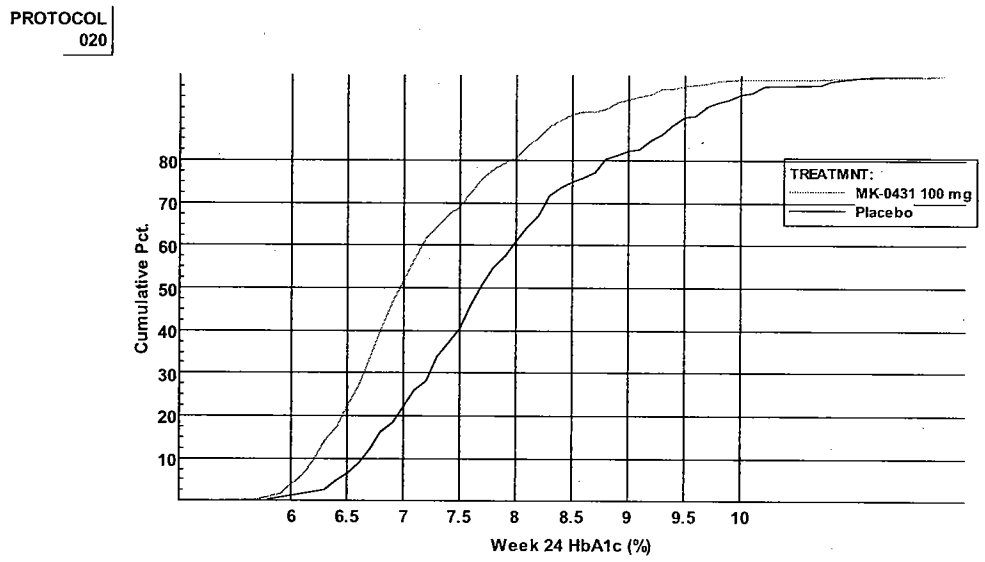
Figure 29 HbA_{1c} change from baseline over time by # of prior medication



APPEARS THIS WAY
ON ORIGINAL

The percentages of patients with endpoint HbA_{1c} 7.0 were 22% (50/224) and 52% (234/453) and the percentages of patients with endpoint HbA_{1c} 6.5 were 6% (14/224) and 22% (101/453) in the placebo group and the MK-0431 100 mg group, respectively (Fig. 30).

Figure 30 Cumulative distribution functions: endpoint HbA_{1c} (%)



Approximately 63% of the patients were on metformin doses of 2000 mg or more per day. Figure 31 displays the mean HbA_{1c} change from baseline during study by metformin use <2000 mg or >=2000 mg. Figure 32 displays HbA_{1c} change from baseline over time by prior medication strata and metformin dose.

Figure 31 HbA_{1c} change from baseline over time by metformin dose

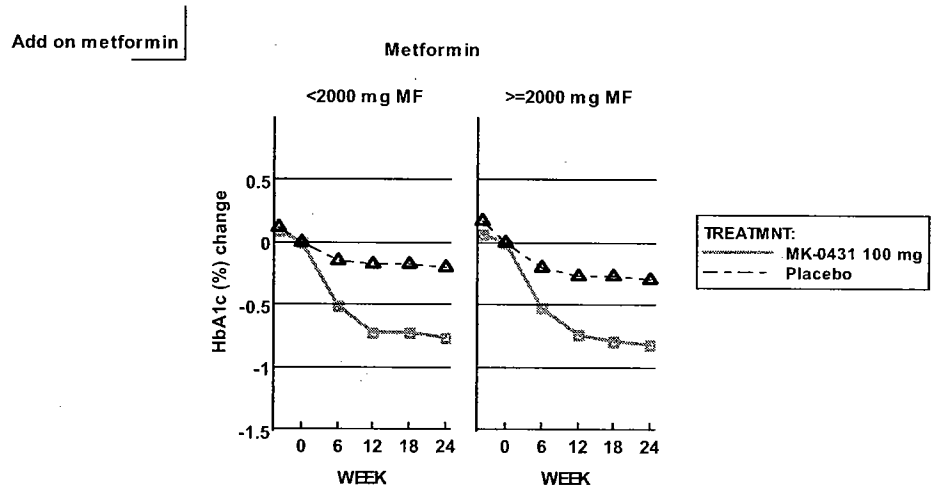
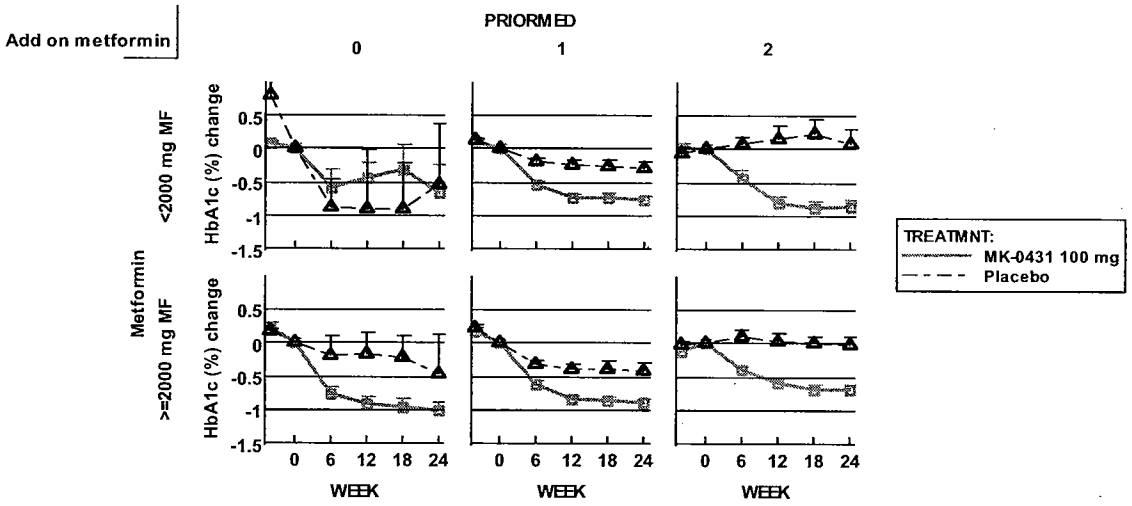
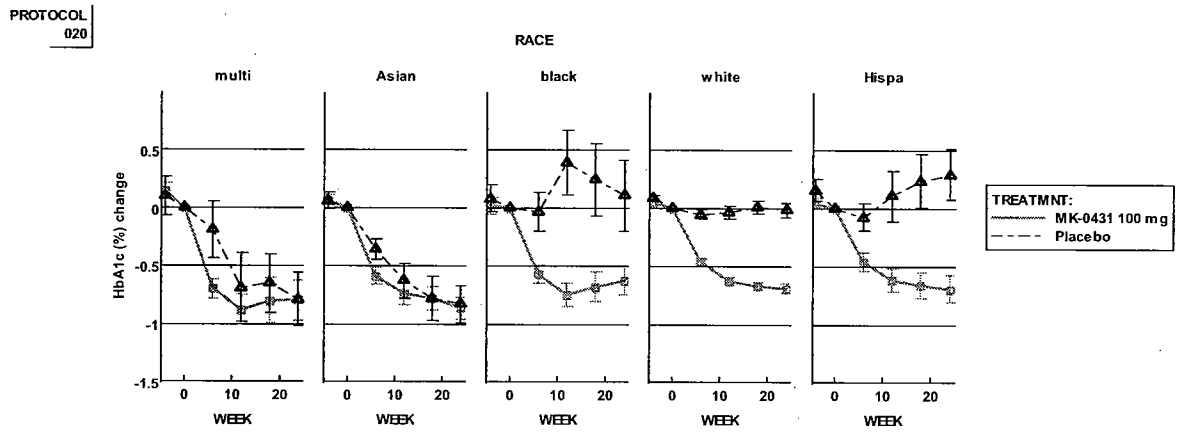


Figure 32 HbA_{1c} change from baseline over time by # of prior AHA and metformin dose



Treatment-by-race interaction was significant ($p<0.01$). The LSM difference in HbA_{1c} change from baseline to 24 weeks between placebo and MK-0431 100 mg was -0.11% for multi race, -0.15% for Asians, -0.65% for Blacks, -0.69% for Whites and -1.0% for Hispanics (Fig. 33).

Figure 33 HbA_{1c} change over time by race



3.1.5 Phase 2 studies and renal insufficiency study (12-week)

Study P010 was a placebo and active (glipizide) controlled dose ranging study. The MK-0431 doses were twice daily 5, 12.5, 25, and 50 mg. Table 33 displays the sponsor's detailed analysis in HbA_{1c} change from baseline at Week 12.

In discussion of a comparison of the 50 and 25-mg bid groups, the sponsor stated that, 'While the 50-mg b.i.d. group was the maximally effective dose across all MK-0431 doses, the difference in HbA1c reduction between the 50- and 25-mg b.i.d. groups (-0.03% with 95% CI of (-0.23, 0.17)) was numerically smaller than observed with the MITT results (-0.11% with 95% CI of (-0.30, 0.08)'.

Therefore, there was no clinically important difference between 25 mg bid and 50 mg bid.

Study P014 was a placebo controlled dose ranging study using MK-0431 once daily doses 25, 50, and 100 mg and twice daily dose 50 mg. Table 34 displays the sponsor's detailed analysis in HbA_{1c} change from baseline at Week 12.

The LSM in HbA_{1c} change from baseline were 0.12, -0.28, -0.44, -0.44 and -0.43 in the placebo, MK-0431 25 mg qd, 50 mg qd, 100 mg qd and 50 mg bid, respectively. The LSM differences from placebo were -0.55 and -0.56 with the same C.I. (-0.76, -0.36) for the 50 mg qd and 100 mg qd, respectively. Therefore, there was essentially no difference between the 50 mg qd and the 100 mg qd.

Figure 34 displays the between group LSM differences and 95% confidence interval for the two 12-week Phase 2 studies. In Study P010 dose response were not well separated for the bid doses above 5 mg and were not well separated above 25 mg qd in Study P014.

Figure 34 LSM Difference from Placebo (95% CI) – Phase 2 Studies

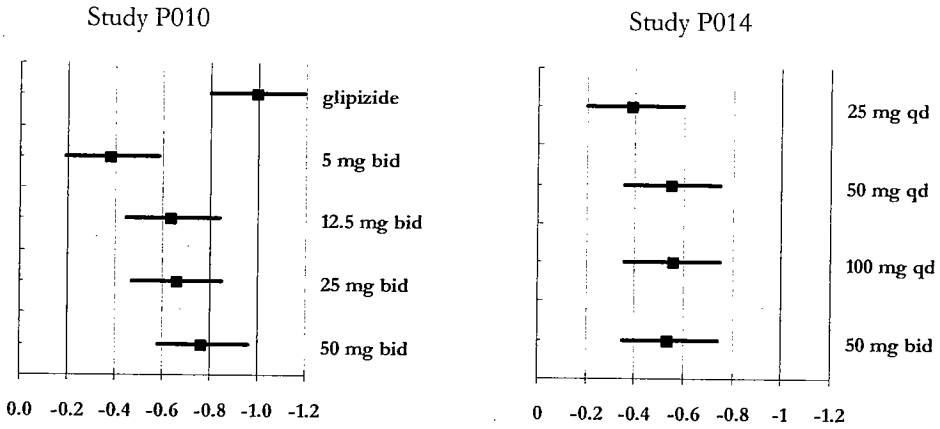


Table 33 Analysis of Change from Baseline in Hemoglobin A1c(%) at Week 12 Modified Intention-To-Treat With Data Carried Forward – Study 010

Treatment	N	Mean		Change From Baseline			
		Baseline (SD)	Week 12 (SD)	Mean (SD)	LS Mean	95% CI for LS Mean	Within-Group p-Value
Placebo	121	7.88 (0.96)	8.14 (1.23)	0.27 (0.90)	0.23	(0.10, 0.37)	<0.001
MK-0431 5 mg b.i.d.	122	7.89 (0.94)	7.77 (1.22)	-0.13 (0.82)	-0.15	(-0.29, -0.01)	0.031
MK-0431 12.5 mg b.i.d.	122	7.85 (0.88)	7.48 (0.98)	-0.38 (0.71)	-0.41	(-0.55, -0.27)	<0.001
MK-0431 25 mg b.i.d.	120	7.89 (0.94)	7.50 (1.14)	-0.39 (0.84)	-0.43	(-0.56, -0.29)	<0.001
MK-0431 50 mg b.i.d.	121	7.83 (0.95)	7.34 (1.01)	-0.49 (0.66)	-0.54	(-0.68, -0.40)	<0.001
Glipizide	119	7.82 (0.95)	7.11 (0.91)	-0.72 (0.84)	-0.76	(-0.90, -0.62)	<0.001

Dose Response Among MK-0431 Doses Versus Placebo	
Stepwise Linear Contrast Test Results (Doses Included in the Current Step)	p-Value
Placebo to MK-0431 50 mg b.i.d.	<0.001
Placebo to MK-0431 25 mg b.i.d.	<0.001
Placebo to MK-0431 12.5 mg b.i.d.	<0.001
Placebo to MK-0431 5 mg b.i.d.	<0.001

Pairwise Differences	Difference in LS Means	95% CI for Difference in LS Means
MK-0431 50 mg b.i.d. versus Placebo	-0.77	(-0.96, -0.58)
MK-0431 25 mg b.i.d. versus Placebo	-0.66	(-0.85, -0.47)
MK-0431 12.5 mg b.i.d. versus Placebo	-0.64	(-0.84, -0.45)
MK-0431 5 mg b.i.d. versus Placebo	-0.38	(-0.58, -0.19)

Pairwise Comparisons for Other Pairs			
Pairwise Comparison	Difference in LS Means	95% CI for Difference in LS Means	p-Value
Placebo or MK-0431 versus Glipizide			
Placebo versus Glipizide	1.00	(0.80, 1.19)	<0.001
MK-0431 5 mg b.i.d. versus Glipizide	0.61	(0.42, 0.81)	<0.001
MK-0431 12.5 mg b.i.d. versus Glipizide	0.35	(0.16, 0.55)	<0.001
MK-0431 25 mg b.i.d. versus Glipizide	0.34	(0.14, 0.53)	<0.001
MK-0431 50 mg b.i.d. versus Glipizide	0.23	(0.03, 0.42)	0.023
Between MK-0431 Doses			
MK-0431 50 mg b.i.d. versus MK-0431 5 mg b.i.d.	-0.39	(-0.58, -0.19)	<0.001
MK-0431 25 mg b.i.d. versus MK-0431 5 mg b.i.d.	-0.28	(-0.47, -0.08)	0.005
MK-0431 12.5 mg b.i.d. versus MK-0431 5 mg b.i.d.	-0.26	(-0.45, -0.07)	0.008
MK-0431 50 mg b.i.d. versus MK-0431 12.5 mg b.i.d.	-0.13	(-0.32, 0.07)	0.197
MK-0431 25 mg b.i.d. versus MK-0431 12.5 mg b.i.d.	-0.02	(-0.21, 0.18)	0.872
MK-0431 50 mg b.i.d. versus MK-0431 25 mg b.i.d.	-0.11	(-0.30, 0.08)	0.261

P-Value for Effect	
Baseline	<0.001
Treatment	<0.001
Prior Anti-hyperglycemic Medication	<0.001

Root Mean Square Error of Change = 0.77

CI = Confidence Interval; LS = Least Squares; SD = Standard Deviation.

From sponsor Table 7-1, Reference P010

Table 34 Analysis of Change from Baseline in Hemoglobin A1c(%) at Week 12 Modified Intention-To-Treat With Data Carried Forward – Study 014

Treatment	N	Mean		Change from Baseline			
		Baseline (SD)	Week 12 (SD)	Mean (SD)	LS Mean	95% CI for LS Mean	Within-Group p-Value
Placebo	107	7.59 (0.89)	7.76 (1.11)	0.17 (0.60)	0.12	(-0.02, 0.26)	0.102
MK-0431 25 mg q.d.	107	7.71 (0.91)	7.47 (1.30)	-0.23 (0.87)	-0.28	(-0.42, -0.14)	<0.001
MK-0431 50 mg q.d.	107	7.60 (0.94)	7.22 (1.02)	-0.38 (0.68)	-0.44	(-0.58, -0.30)	<0.001
MK-0431 100 mg q.d.	106	7.78 (0.90)	7.38 (1.11)	-0.40 (0.81)	-0.44	(-0.58, -0.30)	<0.001
MK-0431 50 mg b.i.d.	108	7.79 (0.85)	7.41 (1.10)	-0.38 (0.76)	-0.43	(-0.56, -0.29)	<0.001

Dose Response Among MK-0431 Once-Daily Doses Versus Placebo			
Stepwise Linear Contrast Test Results (Once-Daily Doses Included in the Current Step)			p-Value
Placebo to MK-0431 100 mg q.d.			<0.001
Placebo to MK-0431 50 mg q.d.			<0.001
Placebo to MK-0431 25 mg q.d.			<0.001
Pairwise Differences	Difference in LS Means	95% CI for Difference in LS Means	
MK-0431 100 mg q.d. versus Placebo	-0.56	(-0.75, -0.36)	
MK-0431 50 mg q.d. versus Placebo	-0.55	(-0.75, -0.36)	
MK-0431 25 mg q.d. versus Placebo	-0.39	(-0.59, -0.20)	
Pairwise Comparisons for Other Pairs			
Pairwise Comparison	Difference in LS Means	95% CI for Difference in LS Means	p-Value
Placebo or MK-0431 Once-Daily Doses versus MK-0431 50 mg b.i.d.			
Placebo versus MK-0431 50 mg b.i.d.	0.54	(0.35, 0.74)	<0.001
MK-0431 25 mg q.d. versus MK-0431 50 mg b.i.d.	0.15	(-0.05, 0.35)	0.132
MK-0431 50 mg q.d. versus MK-0431 50 mg b.i.d.	-0.01	(-0.21, 0.18)	0.910
MK-0431 100 mg q.d. versus MK-0431 50 mg b.i.d.	-0.01	(-0.21, 0.18)	0.900
Between MK-0431 Once-Daily Doses			
MK-0431 100 mg q.d. versus MK-0431 25 mg q.d.	-0.16	(-0.36, 0.03)	0.104
MK-0431 50 mg q.d. versus MK-0431 25 mg q.d.	-0.16	(-0.36, 0.03)	0.106
MK-0431 100 mg q.d. versus MK-0431 50 mg q.d.	-0.00	(-0.20, 0.20)	0.990
p-Value for Effect			
Baseline			0.101
Treatment			<0.001
Prior Anti-hyperglycemic Medication			<0.001
Root Mean Square Error of Change = 0.73			
CI=Confidence Interval; LS=Least Squared; SD=Standard Deviation.			

From sponsor Table 7-1, Reference P014

Study P028 tested MK0-0431 25 mg qd and 50 mg qd in patients with type 2 diabetes and chronic renal insufficiency. There were no efficacy hypothesis in this study but the HbA_{1c} was analyzed to assess glycemic control. The primary analysis was based on the all-patients-as-treated population with no imputation for missing data and data were treated as missing after the initiation of rescue therapy. Table 35 displays results of HbA_{1c} change from baseline.

Table 35 Analysis of Change from Baseline in Hemoglobin A1c(%) at Week 12 Modified Intention-To-Treat With Data Carried Forward – Study 028 (Renal Insufficiency)

Treatment	N	Mean (SD)		Change from Baseline	
		Baseline	Week 12	Mean (SE)	95% CI for Mean
MK-0431	55	7.60 (0.95)	7.01 (0.83)	-0.59 (0.08)	(-0.76, -0.42)
Placebo	25	7.81 (0.90)	7.63 (1.05)	-0.18 (0.13)	(-0.44, 0.08)
Between Treatment Difference				Difference in Means (95% CI)	
MK-0431 vs. Placebo				-0.41 (-0.71, -0.11)	
CI=Confidence Interval; SD=Standard Deviation; SE=Standard Error.					

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Efficacy was consistent in gender and age group for the 100 mg MK-0431 group and the placebo group (p>0.1) (Figs 35 & 36).

Figure 35 Mean HbA_{1c} change from baseline (95% C.I.) by treatment and gender

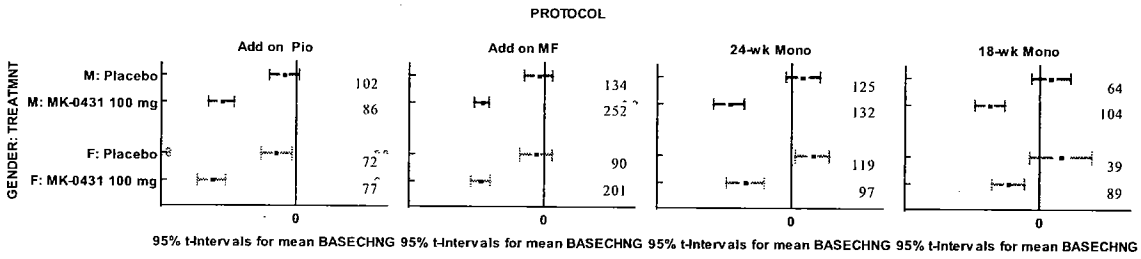
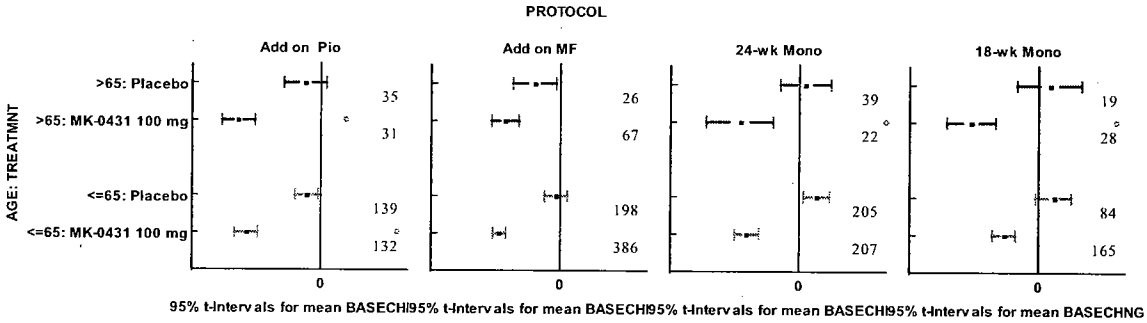
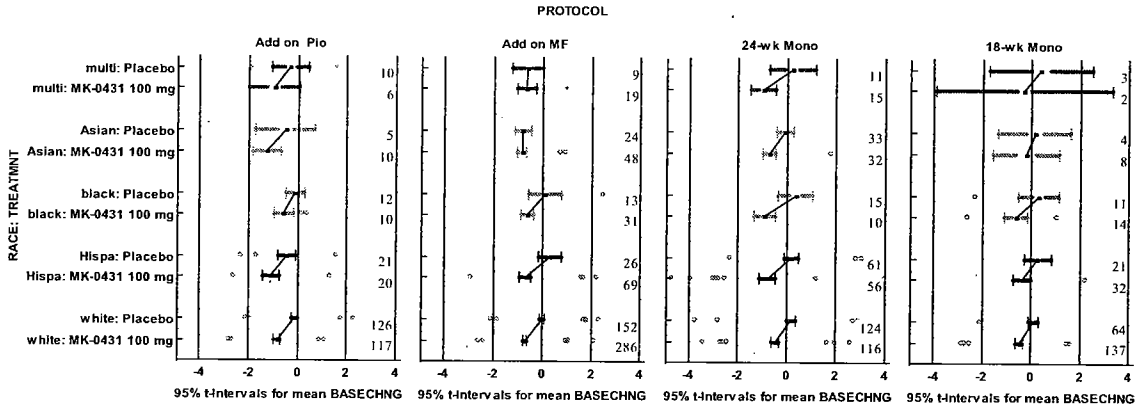


Figure 36 Mean HbA_{1c} change from baseline (95% C.I.) by treatment and age group (>65, ≤65 years)



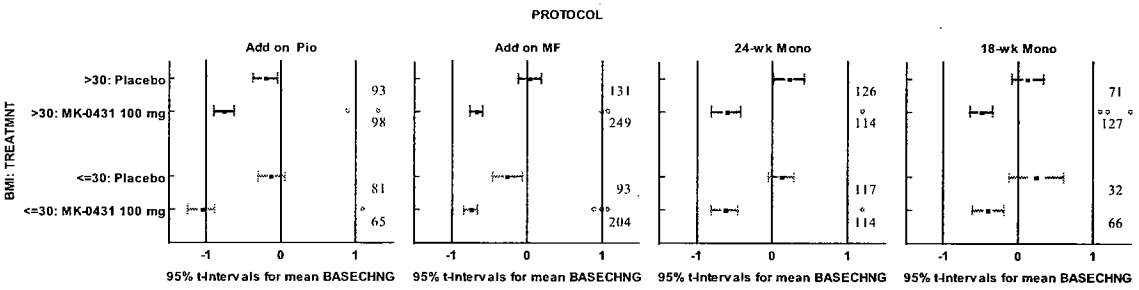
The treatment-by-race interaction was significant for the metformin add on study (p=0.002) (Fig.37).

Figure 37 Mean HbA_{1c} change from baseline (95% C.I.) by treatment and race



The treatment-by-BMI (>30 or ≤30 kg/m²) interaction was significant (p=0.04) for the pioglitazone add on study. The group difference in HbA_{1c} change was -0.89% for BMI ≤30 kg/m² patients and -0.56% for BMI>30 kg/m² patients (Fig 38).

Figure 38 Mean HbA_{1c} change from baseline (95% C.I.) by treatment and BMI ((>30 or ≤30 kg/m²)



Prior medication

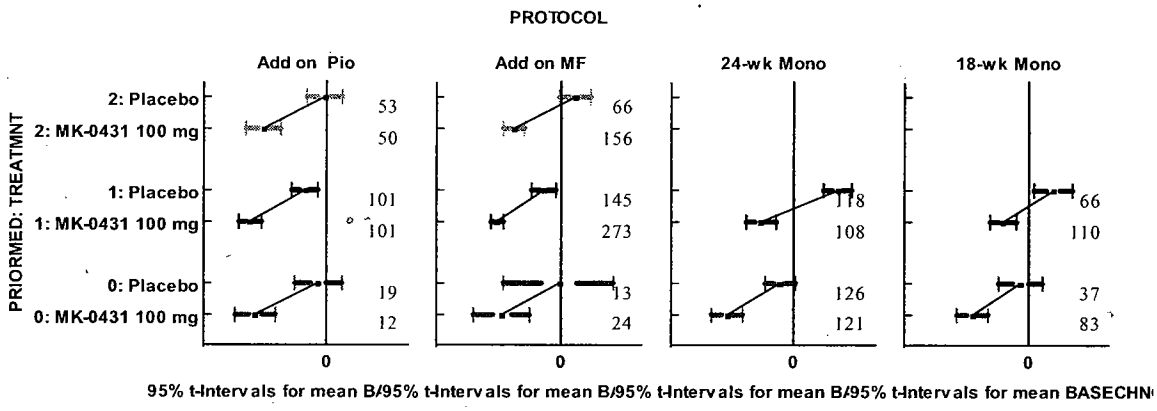
Table 36 displays percentages of patients with prior medication. The 18-week compared to the 24-week monotherapy study had greater percentage of patients on prior AHA (59% vs. 48%).

Table 36 Percentage of patients with prior AHA use - HbA_{1c} APT

	Pio add on		Metformin add on		24-week monotherapy			18-week monotherapy		
	Placebo	100 mg	Placebo	100 mg	Placebo	100 mg	200 mg	Placebo	100 mg	200 mg
N	173	163	224	453	244	229	238	103	193	199
no	19 (11%)	12 (7%)	13 (6%)	24 (5%)	126 (52%)	121 (53%)	121 (51%)	37 (36%)	83 (43%)	83 (42%)
yes	101 (58%)	101 (62%)	145 (65%)	273 (60%)	118 (48%)	108 (47%)	117 (49%)	66 (64%)	110 (57%)	116 (58%)
2 with 1 background	53 (31%)	50 (31%)	66 (29%)	156 (34%)						

The treatment-by-prior medication use was significant (p=0.05) for the 24-week monotherapy study. Comparing placebo patients who washed out their prior AHA to patients not on prior AHA, HbA_{1c} increased from baseline in washed out patients in both monotherapy studies.

Figure 39 Mean HbA_{1c} change from baseline (95% C.I.) by treatment and # of prior medication



HbA_{1c} Category

Figures 40 and 41 display the percentages of patients achieving the glycemic goal of HbA_{1c} <7% and HbA_{1c} <6.5%, respectively, for the APT population.

Figure 40 Percent of patients with HbA_{1c}<7%

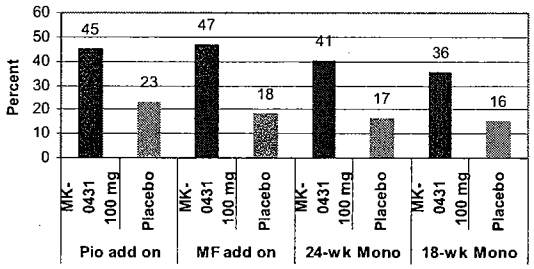


Figure 41 Percent of patients with HbA_{1c}<6.5%

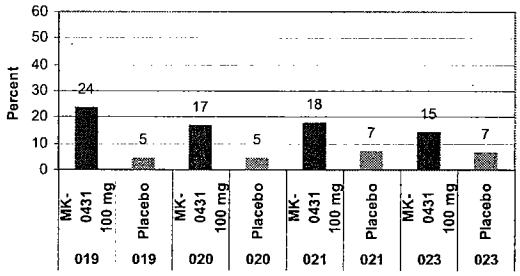
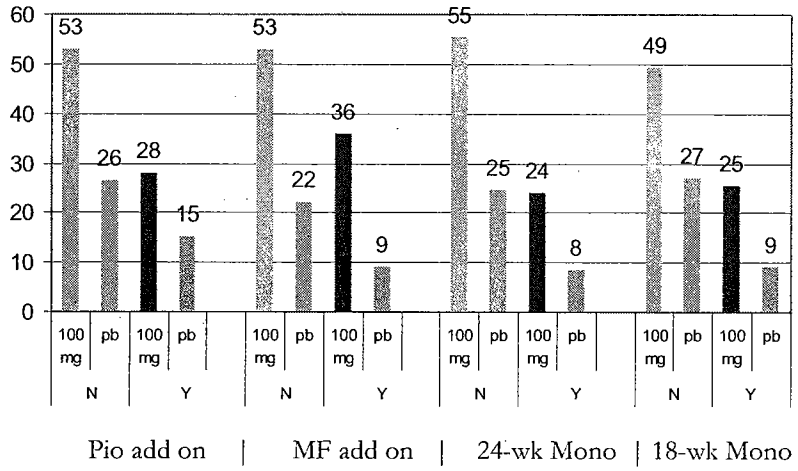


Figure 42 displays the percentages of patients with HbA_{1c}<7% at the end of study by prior AHA use. The percent of patients achieving the 7% glycemic goal was less in the 'yes wash out' patients than 'no wash out' prior AHA patients.

Figure 42 Percent of patients with HbA_{1c}<7% by Prior AHAs 'washed out' (Y/N) -APT



Rescue medication

The monotherapy studies had greater rescue rates than the add-on studies. The placebo rescue rate in the 24-week monotherapy was 21% in the randomized patient population (Table 37). Table 38 displays the percentage of patients rescued in APT (All Patient Treated) population. The rescued patients in the monotherapy studies were mostly patients washed out of their prior AHAs (Table 39). Figures 43 and 44 display the percentage of rescued patients over time using Kaplan Meier survival plot and time in days of each rescued patient, respectively.

Table 37 Percentage of patients rescued – Randomized patients

Pio add on		MF add on		24 wk Monotherapy			18 wk Monotherapy		
Plb	100 mg	Plb	100 mg	Plb	100 mg	200 mg	Plb	100 mg	200 mg
14% (25/178)	6.9% (12/175)	13.5% (32/237)	4.5% (21/464)	20.6% (52/253)	8.9% (21/237)	4.8% (12/250)	18.2% (20/110)	9.3% (19/205)	11.2% (23/206)

Table 38 Percentage of patients rescued - HbA_{1c} APT

Pio add on		MF add on		24 wk Monotherapy			18 wk Monotherapy		
Plb	100 mg	Plb	100 mg	Plb	100 mg	200 mg	Plb	100 mg	200 mg

11.5% (20/174)	6.8% (11/163)	9.4% (21/224)	3.5% (16/453)	15.6% (37/244)	6.1% (14/229)	3.8% (9/238)	12.6% (13/103)	6.7% (13/193)	8.0% (16/199)

Table 39 Percentage of patients rescued in prior treated 'wash out' patients - HbA_{1c} APT

	24 wk Monotherapy			18 wk Monotherapy		
# patients washed out prior AHAs	N=118	n=108	N=117	N=66	N=110	N=116
% (# rescued in prior AHAs/total treatment #)	9.0% (22/244)	5.2% (12/229)	2.1% (5/238)	10.7% (11/103)	4.7% (9/193)	5.5% (11/199)

Figure 43 Percent rescued by time from Kaplan Meier Curve

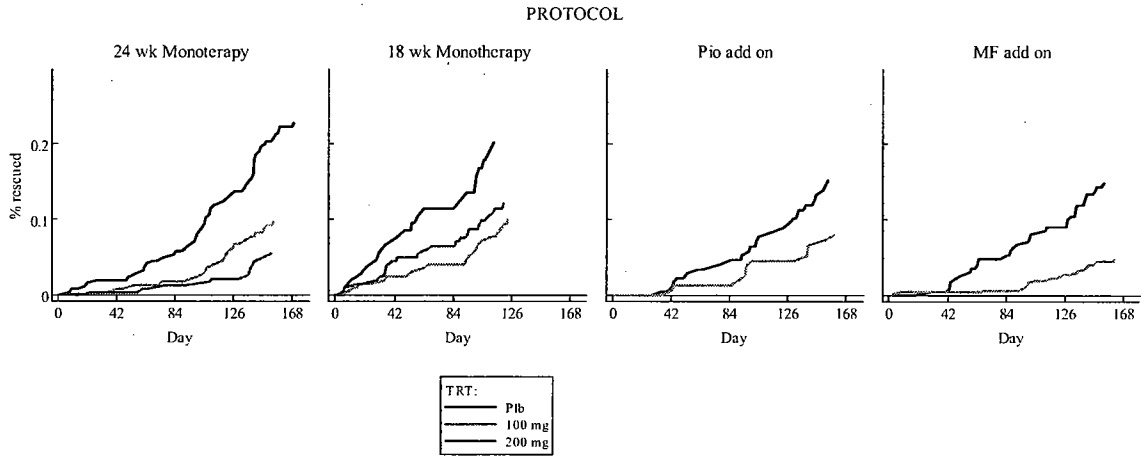


Figure 44 Rescued patient time line by treatment group

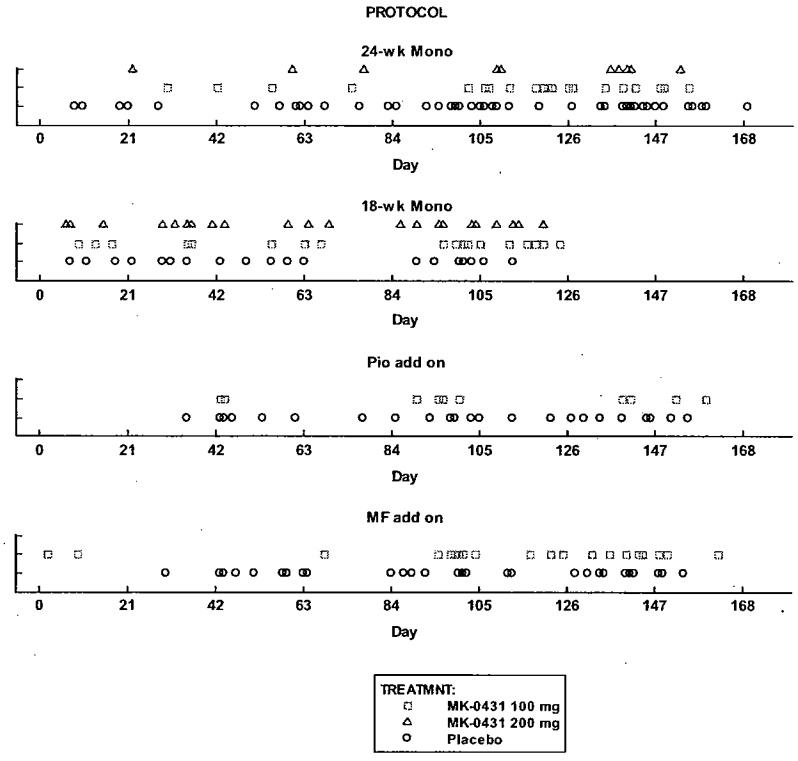


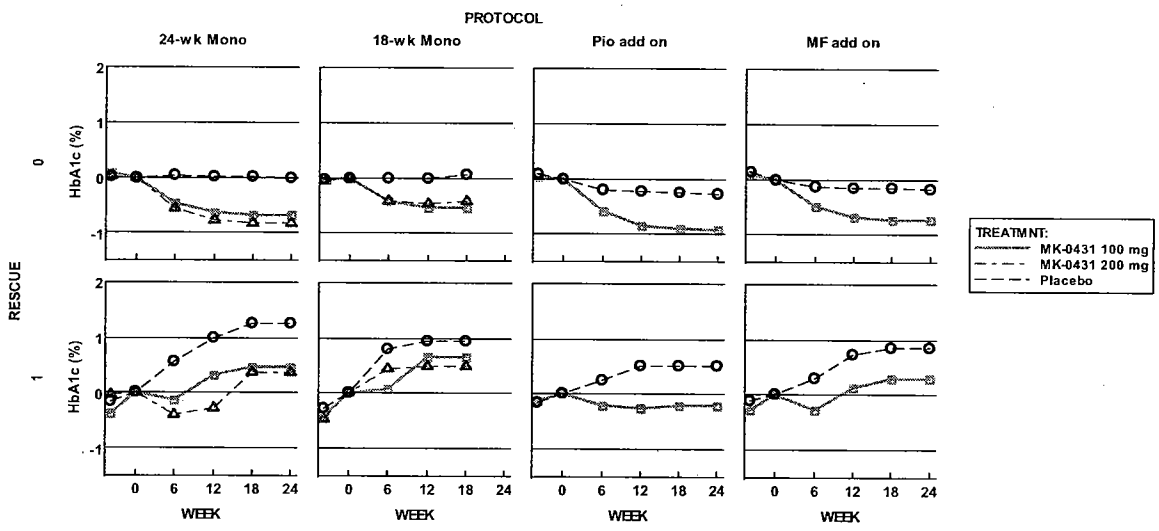
Table 39 and Fig. 45 display the descriptive statistics of mean HbA_{1c} by rescue status.

Table 40 Descriptive statistics in mean HbA_{1c} (SD) by rescue status

Study	Rescue	Placebo				100 mg			
		N	Bl (SD)	Endpoint (SD)	Change (SD)	N	Bl (SD)	Endpoint (SD)	Change (SD)
Pio Add on	0	154	7.89(0.77)	7.63(0.93)	-0.27(0.73)	152	7.97(0.76)	7.04(0.77)	-0.93(0.67)
	1	20	8.83(0.85)	9.34(1.16)	0.51(0.84)	11	9.15(0.62)	8.95(0.82)	-0.21(0.74)
MF Add on	0	203	7.96(0.79)	7.79(0.99)	-0.18(0.86)	437	7.92(0.78)	7.18(0.88)	-0.74(0.69)
	1	21	8.68(0.86)	9.52(0.87)	0.84(0.70)	16	8.99(0.84)	9.28(1.04)	0.29(0.94)
24 wk Mono	0	207	7.94(0.76)	7.92(1.19)	-0.02(0.98)	215	7.96(0.87)	7.27(1.05)	-0.69(0.98)
	1	37	8.57(0.90)	9.81(1.19)	1.24(0.82)	14	8.78(0.84)	9.25(0.94)	0.47(0.99)
18 wk Mono	0	90	7.94(0.86)	7.99(1.24)	0.05(0.88)	180	7.95(0.76)	7.41(0.98)	-0.54(0.79)
	1	13	8.82(0.78)	9.75(1.05)	0.93(0.94)	13	9.24(0.74)	9.9(0.72)	0.66(0.94)

		200 mg		
Study	Rescue	Bl (SD)	Endpoint (SD)	Change (SD)
24 wk Mono	0	8.04(0.91)	7.22(1.06)	-0.82(0.89)
	1	9.06(1.10)	9.42(1.04)	0.37(0.59)
18 wk Mono	0	8.07(0.89)	7.67(1.24)	-0.41(0.93)
	1	8.95(0.73)	9.44(1.00)	0.49(0.87)

Figure 45 mean HbA_{1c} change from baseline by rescue status



5. SAFETY

For nausea, the sponsor reported in a pooled Phase 3 population '... the adverse experience of nausea in the MK-0431 200-mg group occurred at a statistically significantly higher rate than that observed in the non-exposed group. This adverse experience was reported with an incidence of 1.4% (15 patients), 2.9% (13 patients), and 0.6% (5 patients), in the MK-0431 100-mg, 200-mg, and non-exposed patients, respectively.' Table 41 is from Table 2.7.4:62 (Summary of Clinical Safety).

Table 41 Analysis of Gastrointestinal Adverse Experiences
Pooled Phase III Population (P019, P020, P021, P023)
Nausea

Treatment	N	n (%) Patients with at Least One Event	Patient-Years	n/100 pt-yrs
MK-0431 100 mg Exposed	1082	15 (1.4)	452.66	3.31
MK-0431 200 mg Exposed	456	13 (2.9)	174.03	7.47
Non-Exposed	778	5 (0.6)	310.48	1.61

Between Treatment Comparisons Based on Combined Data

Comparison with Non-Exposed	Difference in Proportions† (%) (95% CI†)	p-Value‡
MK-0431 100 mg Exposed vs. Non-Exposed	0.8 (-0.1, 1.7)	0.103
MK-0431 200 mg Exposed vs. Non-Exposed	2.3 (0.5, 4.2)	0.019

Between Dose Comparison	Difference in Proportions† (%) (95% CI†)
MK-0431 200 mg Exposed vs. MK-0431 100 mg Exposed	1.3 (-0.7, 3.2)

CI = Confidence Interval.

† Computed using Cochran-Mantel-Haenszel weighted normal approximation method.

‡ Based on Cochran-Mantel-Haenszel test, with stratification by study.

6. SUMMARY AND CONCLUSIONS

Conclusions and Recommendations

The Phase 2 studies (010 and 014) were well designed with sufficient power on the primary endpoint, HbA1c change from baseline, to assess dose response.

MK-0431 doses studied in Phase 2 ranged from 10 mg to 100 mg daily doses given once a day (qd) or as split doses (bid). 100 mg and 200 mg qd (the latter as monotherapy only) were studied in the Phase 3 trials in diabetic patients with normal renal function. The Phase 3 renal safety study assessed 25 mg and 50 mg qd. The sponsor has proposed a (single) MK-0431 dose of 100 mg qd in patients with Type 2 diabetes and normal renal function and reduced doses of 25 or 50 mg qd in patients with Type 2 diabetes complicated by renal insufficiency.

The phase 2 data showed that efficacy did not appear to be substantially affected if the drug was given as a single daily dose or as split (bid) doses. The 25, 50 and 100 mg doses were consistently superior to placebo in both Phase 2 studies. Maximal efficacy of the drug was achieved at 50 mg with no clear additional clinical benefit from 100 mg.

The proposed doses of 25 mg and 50 mg qd were shown to be effective in the renal-impaired population (study 028).

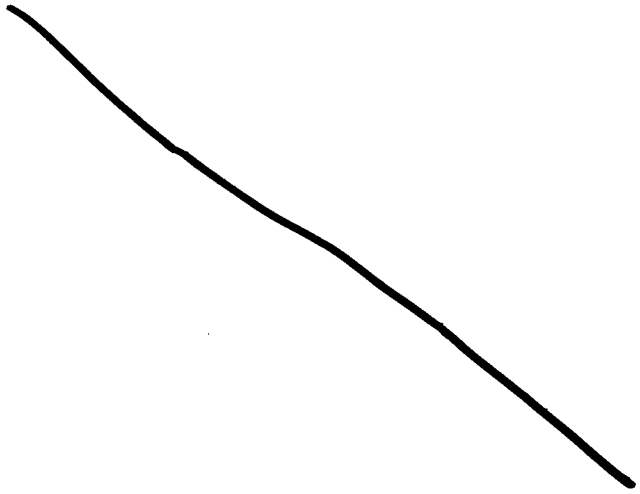
In the four Phase 3 studies, MK-0431 100 mg qd and 200 mg qd were statistically superior to placebo in HbA_{1c} change from baseline in patients with Type 2 diabetes. In the 24-week monotherapy study, the Least Squared Mean (LSM) differences from placebo (95% confidence intervals) in HbA_{1c} change from baseline were -0.79% (-0.96, -0.62) and -0.94% (-1.11, -0.77), respectively, for the 100 mg and 200 mg doses, respectively. In the 18-week monotherapy study, the differences were -0.60% (-0.82, -0.39) and -0.48% (-0.70, -0.26), respectively. Therefore the efficacy of the two doses overlapped in the two monotherapy studies. In the pioglitazone and the metformin add-on studies, the LSM differences

between 100 mg and placebo in HbA_{1c} change from baseline were -0.70 (-0.85, -0.54), and -0.65 (-0.77, -0.53), respectively (Fig. 1). At the End-of-Phase 2 meeting, the Agency had questioned the sponsor's proposed MK-0431 daily dose of 200 mg in the Phase 3 studies. In retrospect, the MK-0431 daily 50 mg dose should have been included in the Phase 3 studies in patients with normal renal function.

In summary, 100 mg was shown to be efficacious as add-on therapy to metformin or pioglitazone. I recommend that, based on the Phase 2 efficacy results showing no clear lessening of clinical benefit for 50 mg compared to the proposed 100 mg dose, daily doses of 50 mg and 100 mg should both be made available to patients with Type 2 diabetes and normal renal function as monotherapy. For diabetic patients with renal insufficiency, the efficacy data suggest that doses of 25 mg and 50 mg are efficacious.

7. LABELING COMMENTS

1.
2.
3.
4.
5.
6.
7.
8.



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

/s/

Lee-Ping Pian
9/15/2006 04:18:43 PM
BIOMETRICS

Todd Sahlroot
9/18/2006 10:25:44 AM
BIOMETRICS

Thomas Permutt
9/18/2006 10:30:47 AM
BIOMETRICS
concur

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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	21,995 / Serial 000
Drug Name:	Januvia™ (Sitagliptin Phosphate)
Indication(s):	An adjunct to diet and exercise to improve glycemic control in patients with type 2 diabetes mellitus as monotherapy and as combination therapy with metformin or a PPAR-y agonist (e,g,, thiazolidinedione) when diet and exercise plus the simgle agent do not provide adequate glycemic control.
Applicant:	Merck & Co., Inc.
Date(s):	Submitted 12/16/05
Review Priority:	Standard
Biometrics Division:	Division 6, HFD-705
Statistical Reviewer:	Steve Thomson, HFD-705
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D., HFD-705
Medical Division:	Metabolsm and Endocrinology Products, HFD-510
Toxicologist:	Reviewer: M. Todd Bourcier, Ph.D., HFD-510 Team Leader: Jeri el-Hage, Ph.D., HFD-510
Project Manager:	Lina Al Juburi, HFD-510
Keywords:	Bayesian analysis, Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY

The Sponsor states that Sitaglipin Phosphate (JANUVIA™), also labeled as MK-0431 in the Sponsor's reports, "is a member of a new class of drugs, the dipeptidyl peptidase IV (DDP-IV) inhibitors." This submission was intended to assess the carcinogenic potential of daily administration of Januvia when administered orally to mice and rats for a period of two years.

1.1. Conclusions and Recommendations

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

Study TT #03-615-0,-2: One-Hundred-Five-Week Oral Carcinogenicity Study in Mice,

and,

Study TT #03-097-0: One-Hundred-Six-Week Oral Carcinogenicity Study in Rats.

In the mouse study there were six treatment groups (i.e., two nominally identically treated control groups, and four dosages of formulation MK-0431 - 50, 125, 250, and 500 mg/kg/day), labeled as Control 1, Control 2, Low, Medium, Medium-High, and High doses respectively. In the rat study, there were five treatment groups (i.e., two supposedly identically treated control groups, and three dosages of formulation MK-0431 - 50, 150, and 500 mg/kg/day), similarly labeled as Control 1, Control 2, Low, Medium, and High doses, respectively. In each study, each treatment group included 50 animals. Vehicle was 0.5% (w/v) methylcellulose 5m/M hydrochloric acid in deoinized water. In both studies treatment was administered orally by gavage for approximately 104 to 106 weeks.

For analyses, the control groups were pooled. Until near the end of the study the high dose group in male mice had the highest mortality rate, but there is no particular dose related trend in the other dose groups and controls (please see Appendix 1 for details). However the tests of homogeneity in survival were not statistically significant, consistent with the hypothesis of homogeneity in survival (Males: Logrank p = 0.7095 & Wilcoxon p = 0.6318). Treatment group related differences in mortality were even less apparent in female mice and female rats (Female mice: Logrank p = 0.5553 & Wilcoxon p = 0.3735, Female rats: Logrank p = 0.7442 & Wilcoxon p = 0.5624). In male rats the high dose group clearly had the highest mortality, while the survival curves in the low and middle dose groups almost coincided, and both completely dominated the survival curves of the high dose group. These apparent differences were highly statistically significant (Male rats: Logrank p = 0.0012 & Wilcoxon p = 0.0009). In male rats the test of trend in dose was also highly statistically significant (p = 0.0002). The Bayesian assessments of mortality seemed to be equivocal. Like the usual frequentist tests, the deviance information criteria (DIC) for the assessments in mice and in female rats suggested that there was no difference between treatment groups in mortality. However for male rats the model with no treatment group differences had lower DIC than the model with trend in dose or

differences in treatment groups. Note that the model with a high dose group and homogeneous effects otherwise had a slightly lower DIC, and the posterior distribution of the parameter corresponding to a high dose group was clearly bounded away from zero (please see Appendix 2 for details).

To adjust for the multiplicity of comparisons involved in a tumorigenicity analysis for standard rodent models the Agency analysis follows the Haseman-Lin-Rahman rules described in Section 1.3.1 below. Even without adjusting for multiplicity, there were no statistically significant trends and differences in tumorigenicity in either mouse gender (please see Appendix 4 for details). In rats the pattern of statistical significance was more complicated. There was statistically significant evidence of tumorigenicity in liver tumors in both genders and adrenal cortex tumors in males. Using the incidence in the pooled controls, except for carcinoma in females, the liver tumors would be classified as common tumors and the remaining as rare tumors. Following the Haseman-Lin-Rahman rules to adjust for multiplicity, for an overall 10% or so error rate, in female rats both the trend tests and the pairwise comparisons between the high dose and the pooled controls were statistically significant for both carcinoma and pooled adenoma/carcinoma of the liver (all unadjusted, observed $p \leq 0.0041$). Similarly the trend tests were statistically significant for adenoma, carcinoma, and pooled adenoma/carcinoma in the liver of male rats (all observed $p \leq 0.0044$). For male rats the pairwise comparisons between the high dose group and the pooled controls were statistically significant for carcinoma and pooled adenoma/carcinoma of the liver (both $p \leq 0.0050$). For adenomas in the liver of male rats the pairwise comparison between the high dose group and the pooled controls was close to statistical significance ($p = 0.0120$). Also, with male rats the trend test in adenomas of the adrenal cortex was statistically significant ($p = 0.0215$), and close to significance for the pooled adenomas/carcinomas ($p = 0.0269$). However the corresponding pairwise tests between the high dose group and the pooled controls in the adrenal cortex were not statistically significant.

1.2. Brief Overview of the Studies

Two studies, both typical rodent studies, were submitted:

Sponsor Study TT #03-615-0,-2: A 105 Week Oral Carcinogenicity Study in Mice

The Sponsor indicates that animals were randomized into six study groups, each with 50 animals per gender. Within each group (i.e., two control groups, and four dosages of formulation MK-0431 - 50, 125, 250, and 500 mg/kg/day), treatment was administered orally by gavage daily for approximately 105 weeks.

Sponsor Study TT #03-097-0: A 106 Week Oral Carcinogenicity Study in Rats

The Sponsor indicates that animals were randomized into five study groups, each with 50 animals per gender. Within each group (i.e., two vehicle control groups, and three dosages of

formulation MK-0431 - 50, 150, and 500 mg/kg/day), treatment was administered orally by gavage for approximately 106 weeks.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

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

1. Two Control Groups:

Both studies used dual controls. Displays of data, including plots of survival curves and tables of tumor incidence, differentiate between these two controls. However, since any difference between these two groups should be due to factors to be treated as random, for all statistical analyses the control groups were pooled.

2. Survival Analysis:

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

The Sponsor provides a sequential series of logrank tests of homogeneity in survival. The overall test of homogeneity is over the pooled controls, the low, medium, medium-high, and high dose groups. This is followed by a test of homogeneity over the pooled controls, the low, medium, and medium-high dose groups, followed by a test of homogeneity over the pooled controls, the low, and medium dose groups. Finally there is a pairwise test of homoegeneity between the pooled controls and the low dose group. This can be interpreted as a search in the ordered treatment groups for treatments adjacent in dose with similar patterns of survivability, and thus as a test for general trend in survival over dose. Although not suggested by the Sponsor, note that these could have been cast as closed tests, and thus would not need an adjustment for multiplicity.

3. Tests in Neoplasms:

The Sponsor states that all animals in all dose groups were exhaustively checked by the pathologist for tumors. Differences in tumor incidence rates between dose groups were analyzed

using the methods of Peto, *et al*, (1980), which adjusts for time of death and cause of death. Note that the Peto tumorigenicity analyses were conducted using the FDA program supported by Dr. Ted Guo.

4. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involves a large number of statistical tests, which in turn necessitates an adjustment in experiment-wise Type I error. Current FDA practice is based on the Haseman-Lin-Rahman rules. Namely, based on his extensive experience with such analyses, for pairwise tests comparing control to the high dose group, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Based on simulations and their experience, Lin & Rahman (1998) proposed a p-value adjustment for tests of trend. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of trend, rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level. In this analysis we will use the observed incidence in the pooled vehicle groups to decide if a tumor is rare or common.

The Sponsor's analysis apparently uses Sidak's inequality to adjust for the multiplicity. If one is most strongly interested in controlling Type I error (i.e., the error of concluding there was evidence of a dose relation to tumorgenicity when there was no such relation), then this is appropriate. The Agency 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).

5. Validity of the Designs:

Lin and Ali (1994), quoting work by Haseman, have suggested that a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. From the survival plots in the Appendix, it is evident that this value was clearly superceded in both studies, and in fact was exceeded by the number of animals that survived to the terminal sacrifice. This may indicate adequate exposure in both studies, but such a conclusion requires the expertise of the toxicologist.

Traditionally, in analyses performed in the United States, the highest dose should be close to the Maximum Tolerated Dose (MTD) to achieve the greatest likelihood of tumorigenicity. Chu, Ceuto, and Ward (1981), citing earlier work by Sontag et al. (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' "

The following table displays the mean weights in each treatment group in each study, and the changes in mean weights. This is computed in two ways. First at days 365 and 722 the mean weight changes of those animals that have survived to that time point is presented. The mean of these changes is labeled as the "Per Animal Change". Finally, also at day 722, the differences per treatment group between the group mean weight at baseline and the group mean weight of surviving animals is presented. Note that this difference, under the heading "Group Mean Change", is not the same as the "Per Animal Change", but they should generally be close. Finally the last column shows the percent in group mean change in each treatment group relative to the control group change. More than a 10% body weight gain decrement relative to controls (as is shown in male mice below) may indicate that the high dose is over the MTD. However, this body weight gain decrement is still close to 10%.

Table 1. Summary of Weights and Weight Changes in Dose Groups

	Days on drug at observation								
	Baseline		365		722				
	Weight		Per Animal		Per Animal			Group	Group
	N	Mean	N	Mean	Change	N	Mean	Change	Mean Change	Relative % Change
Male Mice									
Control	100	30.48	95	47.45	17.0	58	45.99	15.8	15.52	-
Low	50	30.15	46	48.23	18.1	30	45.56	15.4	15.44	99%
Medium	50	29.80	46	46.72	17.0	34	44.56	14.8	14.76	95%
Medium-High	50	29.72	48	47.99	18.4	30	44.46	14.9	14.75	95%
High	49	30.76	39	47.06	16.4	26	44.45	14.0	13.69	88%
Female Mice									
Control	100	23.08	96	34.09	11.0	69	35.10	12.1	12.02	-
Low	50	23.13	47	36.70	13.7	31	37.19	14.2	14.07	117%
Medium	50	22.83	48	36.20	13.4	32	37.71	14.9	14.88	124%
Medium-Hih	50	23.02	48	34.99	12.0	30	35.77	12.9	12.75	106%
High	50	22.62	47	33.61	11.0	29	36.07	13.5	13.44	112%
Male Rats									
Control	100	151.06	99	547.55	396.6	78	554.00	403.2	402.9	-
Low	50	151.18	50	537.94	386.8	32	532.06	380.9	380.9	95%
Medium	50	152.74	49	531.94	379.3	33	532.24	380.9	379.5	94%
High	50	149.02	46	508.52	359.3	24	505.92	353.7	356.9	89%
Female Rats									
Control	100	117.64	99	296.84	179.3	61	295.49	177.7	179.2	-
Low	50	118.60	50	293.84	175.2	35	293.31	174.7	175.2	98%
Medium	50	117.96	45	290.13	172.1	29	290.97	173.1	173.0	97%
High	50	119.64	43	291.40	171.0	32	299.84	179.2	180.2	101%

The body weight gain data in Table 2 show that there was a close to 10% body weight gain decrement in the high dose groups in male mice and rats. It may be concluded that the high doses used in these two groups are close to the MTD. However, female mice and rats did not show any body weight gain decrement. That may be an indication that the high doses used in these two groups were under the MTD.

The mortality data of the mouse study in Table 4 below show that both the male high dose group and the female high dose group had slightly higher mortality than the corresponding control groups. The mortality data for the rat study in Table 7 show that the female high dose group had slightly higher mortality than the control group. However, the male high dose group had much higher mortality than the control group. Based on the mortality data, it can be concluded that the high dose is close to the MTD for male mice, female mice, and female rats, and may exceed the MTD for male rats.

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

1.3.2. Statistical Findings

For all statistical analyses, the control groups were pooled. Until near the end of the study, in the male mice the high dose group tended to have the highest mortality, but with no particular dose related trend in the other dose groups and controls (please see Appendix 1 for details). However, as shown in the table below, the omnibus test of treatment differences was not statistically significant ($p = 0.7095$). In female mice and female rats, there seemed to be no consistent differences in mortality among the dose groups, or any other particular dose related pattern in survival. However there were clear differences in survival in male rats. The high dose group consistently had the highest mortality. The low and medium dose groups tracked quite closely, with consistently higher mortality than the pooled controls, but less than the high dose group. For both genders of mice, and also female rats the omnibus tests of homogeneity in survival were not statistically significant. However, differences in survival in male rats were statistically quite significant ($p = 0.0012$), and corresponded to a generally increasing mortality in dose, except that the 50 and 150 mg/kg/day groups were almost coincident.

Table 2. Tests of Homogeneity and Trend in Survival

	Mice			Rats		
Gender	Log Rank	Wilcoxon	Trend	Log Rank	Wilcoxon	Trend
Male	0.7095	0.6318	0.2617	0.0012	0.0009	0.0002
Female	0.5553	0.3735	0.2499	0.7442	0.5624	0.8507

In male and female mice, and in female rats the Bayesian tests of survival seem to be consistent with no particular dose related differences in survival, as is indicated in the classical frequentist tests above. However, unlike the classical frequentist tests above, in male rats in the Bayesian analysis the model with no treatment group differences seemed to fit better than the models with trend in dose or differences in treatment groups. However there was weak evidence

that the model with a high dose effect and otherwise homogeneous effects fit slightly better than the other models (please see Appendix 2 for details).

For tumorigenicity, in both mouse genders, even without adjusting for multiplicity, there were no statistically significant trends and differences between the pooled controls and the high dose groups. In rats the pattern of statistical significance was more complicated. There was statistically significant evidence of tumorigenicity in liver tumors in both genders and adrenal cortex tumors in males. Using the incidence in the pooled controls, except for carcinoma in females, the liver tumors would be classified as common tumors and the remaining as rare tumors. Then by the Haseman-Lin-Rahman rules described in Section 1.3.1 above, in female rats both the trend tests and the pairwise comparisons between the high dose and the pooled controls were statistically significant for both carcinoma and pooled adenoma/carcinoma of the liver (all observed, and unadjusted for multiplicity, $p \leq 0.0041$). Following these same rules, in male rats the trend tests were statistically significant for adenoma, carcinoma, and pooled adenoma/carcinoma in the liver (all $p \leq 0.0044$). The pairwise comparisons between the high dose group and the pooled controls were statistical significant for carcinoma and pooled adenoma/carcinoma of the liver in male rats (both $p \leq 0.0050$). For adenomas in male rats the pairwise comparison between the high dose group and the pooled controls was close to statistical significance ($p = 0.0120$). With male rats the trend tests in the adrenal cortex was statistically significant ($p = 0.0215$), and close to significance for the pooled adenomas/carcinomas ($p = 0.0269$). However the corresponding pairwise tests between the high dose group and the pooled controls in the adrenal cortex were not statistically significant.

The statistically significant differences in the Agency analysis are summarized in Table 3 below. There were differences between the exact p-values computed in the FDA analysis and those provided by the Sponsor. However, after adjusting for multiplicity, either by the Sponsor's methods or by the Agency's, these differences did not seem to have an impact on conclusions. Those comparisons that were statistically significant are denoted by an asterix ("*"). Those comparisons close to statistical significance are denoted by a question mark ("?"). No other comparisons achieved statistical significance.

Table 3. Tests of Trend and Difference Between the High and Control Groups in Tumorigenicity

Male Mice: None
Female Mice: None

	Pooled Controls	Low	Medium	High	p-values Trend	Hi vs Lo
Male Rats:						
Adrenal/Cortex						
Adenoma	0	0	1	2	0.0215*	0.0941
Adenoma/Carcinoma	0	0	2	2	0.0269?	0.0941
Liver						
Adenoma	2	0	1	5	0.0015*	0.0120?
Carcinoma	4	6	1	7	0.0044*	0.0050*
Adenoma/Carcinoma	6	6	2	12	0.0000*	0.0000*

Table 3. (cont.) Tests of Trend and Difference Between the High and Control Groups in Tumorigenicity

Female Rats:						
Liver						
Carcinoma	1	1	1	6	0.0008*	0.0041*
Adenoma/Carcinoma	2	3	4	9	0.0001*	0.0004*

*-Statistically Significant (approximately 10-15%)
?-Close to Statistical Significance
Otherwise not statistically significant

2. INTRODUCTION

2.1. Overview

Results from a study in [redacted]:CD-1 (ICR) BR mice and a study in [redacted]:CD®(SD)IGS BR rats were submitted to assess the carcinogenic potential of Januvia™. Following the usage in the Sponsor's reports this drug is also labeled as MK-0431 in the following discussion.

2.2. Data Sources

For both studies, the following SAS transport data sets were included in the FDA electronic data room (edr):

MICRO.XPT, MORTAL.XPT, TUMOR.XPT, and WEIGHTS.XPT.

These describe the results of the microscopic examination (histopathology), mortality, tumorigencity, and body weights, respectively.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

Results on both studies are presented below.

3.2.1. Study TT #03-615-0,-2: A 105 Week Oral Carcinogenicity Study in Mice

Animals were 39 to 41 days old at the start of the study. Six treatment groups were formed for each gender in [redacted]:CD-1 (ICR) BR mice (50/sex/group). Oral treatment was administered by gavage for approximately 105 weeks. There were three MK-0431 treatment

groups, labeled "Low", "Medium", "Medium-High" and "High", corresponding to treatment with 50, 125, 250, and 500 mg/kg/day of MK-0431 suspended in vehicle. Otherwise identical Control Groups 1 and 2 were treated with vehicle alone. The dosing volume for all animals was presented as 10 mL/kg. An additional 5 animals per sex/group were included "as replacement animals for mice that died or were sacrificed due to causes unrelated to treatment during the first 8 weeks of the study." (page 9 of Sponsor's report)

Animals were caged two to three mice together and were observed daily for mortality and at least once weekly for physical signs. Food and water were available ad libitum. "All animals were palpated for masses, generally once every 4 weeks starting in Drug Week 26, to provide information regarding the onset of possible neoplasms. Body weights were recorded pretest, once in Drug Week 1, twice per week through Drug Week 13, and once per week thereafter. . . . Complete necropsies, including examination and collection of tissues from an extensive list, were done on all animals." (page 9, Sponsor's report) The study was conducted at a Merck laboratory in France from 23 June 2003 to 27 June 2005.

Although not displayed in this review, for both genders, animal weights never showed any particular trend relative to dose. In fact, there were no statistically significant differences between treatment groups in mean measured weights at any time points in the study.

3.2.1.1 Sponsor's Results and Conclusions

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

Survival analysis:

The Sponsor provided the results of log rank tests of homogeneity in survival over the pooled controls (i.e., Control groups 1 and 2 are pooled for the analysis) and the four doses of Januvia, plus the comparisons deleting the high dose group, then both the high dose group and the medium-high dose, then deleting the three highest doses, and finally including only the controls and the low dose group. However no treatment differences in this test were statistically significant (all $p>0.500$, except for the comparison between the low dose group and the pooled controls in female mice, $p = 0.302$).

These mortality results are summarized in the following table. For each treatment group, the number of animals, the number of natural deaths, the percent who survived to the end of the study, and the p-values for the tests of homogeneity are presented. The results under the high dose group correspond to a test of homogeneity over the pooled controls (i.e., 0 dose) through the dosages of the low, medium, medium-high, and high dose groups. The results under the medium-high dose group correspond to a test of homogeneity over the pooled controls and low, medium, and medium-high dose groups. The results under the medium dose group correspond to a test of homogeneity over the pooled controls and the low and medium dose groups. The results under the low dose group correspond to a test between the pooled controls and the low dose

group. When high dose groups are statistically significant and lower doses are not, this can be interpreted as a test of trend.

Table 4. Summary of Mortality in Mice (dose/kg/day)

Males	Control Groups 1 and 2	Low 50 mg	Medium 125 mg	Medium-High 250 mg	High 500 mg
Group size	100	50	50	50	50
# deaths	42	21	17	21	24
Percent Survival	58%	58%	66%	58%	52%
p-value trend		>0.500	>0.500	>0.500	>0.500

Females	Control Groups 1 and 2	Low 50 mg	Medium 125 mg	Medium-High 250 mg	High 500 mg
Group size	100	50	50	50	50
# deaths	33	21	20	20	21
Percent Survival	67%	58%	60%	60%	58%
p-value trend		0.302	>0.500	>0.500	>0.500

Tumorigenicity analysis:

The Sponsor conducted Peto type analyses to compare the incidence of various neoplasms (see Tables A.3.1 and A.3.2 in Appendix 3). The only statistically significant trends in mice were negative in dose (i.e., decreasing trend in tumorigenicity with increasing dose), and do not remain statistically significant after adjusting for multiplicity.

3.2.1.2 FDA Reviewer's Results

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

Survival analysis:

As with the Sponsor's analysis, for both genders there was no overall statistically significant difference in survival among the five treatment groups, including the pooled controls (Males: logrank p = 0.7095 and Wilcoxon p = 0.6318, Females: logrank p = 0.5553 and Wilcoxon p = 0.3735). Strictly speaking, lack of evidence of heterogeneity in survival should not be treated as proof of homogeneity in survival. Nonetheless, it does seem indicative of homogeneity in mortality among the treatment groups.

Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1, along with more details of the analysis. The following tables (Table 5 for male mice, Table 6 for female mice) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and give the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage is the percent survived at the end of the

interval, as estimated using a Kaplan-Meier estimate on the ungrouped data. Note again that mortality results seem to be quite consistent across dose groups.

Table 5. Summary of Male Mice Mortality (dose/kg/day)

Period (Weeks)	Control 1	Control 2	Low 50 mg	Medium 125 mg	Medium-High 250 mg	High 500 mg
1-50	0/50 [1] 100%[2]	4/50 92%	4/50 92%	4/50 92%	2/50 96%	9/50 82%
51-78	6/50 88%	7/46 78%	5/46 82%	3/46 86%	7/48 82%	7/41 68%
79-91	4/44 80%	6/39 66%	8/41 66%	6/43 74 %	6/41 70 %	4/34 60%
92-103,4	7/40 66%	8/33 50%	4/33 58%	3/37 68 %	6/35 58 %	4/30 52%
Terminal	33	25	29	34	29	26

[1] number deaths / number at risk

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

Merely for display, survival in the two control groups is shown separately. For testing the controls are pooled. The tests of homogeneity in survival were based on the pooled control groups. Note there seems to be no statistically significant evidence of differences in survival across treatment groups.

The similar table (Table 6) for females is given below:

Table 6. Summary of Female Mice Mortality (dose/kg/day)

Period (Weeks)	Control 1	Control 2	Low 50 mg	Medium 125 mg	Medium-High 250 mg	High 500 mg
1-50	2/50 [1] 96%[2]	2/50 96%	2/50 96%	1/50 98%	2/50 96%	3/50 94%
51-78	4/48 88%	5/48 86%	5/48 86%	3/49 92%	10/48 76%	7/47 80%
79-91	4/44 80%	4/43 78%	6/43 74%	3/46 86%	4/38 68 %	9/40 68%
92-103,4	6/40 68%	6/39 66%	7/37 60%	13/43 60 %	4/34 60 %	5/31 58%
Terminal	34	33	30	30	30	29

[1] number deaths / number at risk

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

While there seems to be an apparent slightly higher mortality in the higher dose groups, the hypothesis of homogeneity in mortality across treatment groups is not rejected. So, while

absence of proof is not proof of absence we can conclude there is no strong evidence of treatment differences in survival.

Tumorigenicity analysis:

The Peto mortality adjusted tests of trend in the incidence of neoplasms over the four MK-0431 groups and the difference between the pooled controls and the high dose group are displayed in Appendix 4. For 12 or fewer tumors the results of an exact test (assuming fixed marginals) are provided. For more than 12 tumors the results from an asymptotic test are given. For tumorigenicity, in both mouse genders, even without adjusting for multiplicity, there were no statistically significant trends and differences between the pooled controls and the high dose groups.

3.2.2. Study TT #03-097-0: A 106 Week Oral Carcinogenicity Study in Rats

Animals were 38 days old at the start of the study. Five treatment groups were formed with each of fifty male and female [redacted]:CD®(SD)IGS BR rats (i.e., 50/sex/group). Oral treatment was administered once daily by gavage for approximately 106 weeks. The Low, Medium, and High dose groups were treated once daily with 50, 150, or 500 mg/kg/day, respectively, of MK-0431 suspended in vehicle. Otherwise identically treated, Control Groups 1 and 2 were treated with vehicle. The dosing volume for all animals is presented as 5 mL/kg.

Animals were housed individually and checked for mortality daily. The Sponsor reports that female and male rats were provided with 16g/day and 22g/day of [redacted] Water was available ad libitum. Animals were observed once per week for physical signs. "Beginning in Drug Week 26, all animals were palpated for masses every 4 weeks to provide information regarding the onset of possible neoplasms for use in statistical analyses. Body weights were recorded pretest, once in Drug Week 1, twice per week through Drug Weeks 2 to 13, and once per week thereafter. . . . Complete necropsies, including examination and collection of tissues from an extensive list, were done on all animals." (page 8, Sponsor's report) The study was conducted at a Merck laboratory in France from 9 July 2003 to 14 July 2005.

Although not presented in this review, the animal weight pattern is more complicated in the rat species than in mice. For male rats, from the beginning of the study to Day 99, there were no statistically significant differences in mean weights among the various dose groups. In females, for Days 54-99 there were no statistically significant differences between treatment groups in mean weights. At all other times, for both genders, there were statistically significant differences for both genders. In males the low dose group usually had slightly higher mean weights than the pooled vehicle group, but the weights for these groups were generally quite close. However, both the pooled control group and the low dose group had weight means that were slightly higher than those of the higher dose groups. For females, after Day 99, the pooled vehicle control group had higher mean weight than the other dose groups. Further, for both

genders, with the exception noted above, after Day 99, the mean weights almost always showed a decreasing trend relative to increasing dose.

3.2.2.1 Sponsor's Results and Conclusions for Rats

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

Survival analysis:

The following table (Table 7) displays, for each treatment group, the number of animals, the number of natural deaths, the percent who survived to the end of the study, and the p-values for logrank tests of homogeneity over doses. The results under the High dose group correspond to a test of homogeneity over dosages of the pooled controls (i.e., 0 dose) and the four MK-0431 treatment groups. The results under the Medium dose group correspond to a test of homogeneity over the dosages of the pooled controls and the Low, and Medium dose groups. Similarly, the results under the Low dose group are the results of a pairwise comparison of the Low treatment group to the pooled controls, respectively. In male rats there is statistically significant evidence of a lack of homogeneity in survival (p = 0.001). No such pattern is evident in females.

Table 7. Summary of Mortality in Rats (dose/kg/day)

Males	Pooled Control Groups 1 and 2	Low 50 mg	Medium 150 mg	High 500 mg
Group size	100	50	50	50
# deaths	22	18	17	26
Percent Survival	78%	64%	66%	48%
p-value trend		0.056	0.111	**0.001**

Females	Pooled Control Groups 1 and 2	Low 50 mg	Medium 150 mg	High 500 mg
Group size	100	50	50	50
# deaths	39	17	21	18
Percent Survival	55%	56%	74%	50%
p-value trend		>0.500N	>0.500	>0.500N

N denotes negative trend

Tumorigenicity analysis:

The Sponsor also conducted Peto type analyses of tumorigenicity in rats (see Tables A.3.3 and A.3.4 in Appendix 3). In rats the pattern of tumorigenicity was somewhat more complicated than with mice. In male rats there were statistically significant trends in Liver Hepatocellular Adenomas and Carcinomas (unadjusted p = 0.002 and p = 0.006, respectively). Note that these would be classified as common tumors. Adjusting for multiplicity using the Haseman-Lin-Rahman rules or the Sponsor's Sidak adjustment only the trend in Liver Hepatocellular Adenomas remained statistically significant (Sidak adjustment p = 0.016). Using either adjustment, Liver Hepatocellular Carcinomas in male rats remained statistically significant (Sidak adjustment p = 0.066). In males the trend in Islet cell adenoma was (unadjusted)

statistically significant (p = 0.023), but negative in trend in dose, while adenomas in the adrenal cortex were positive in trend, but again, barely statistically significant (p = 0.023). Using either the Haseman-Lin-Rahman rules or the Sidak correction to adjust for the multiplicity in tests, neither trend remained statistically significant. In female rats Hepatocellular Carcinomas would be classed as rare tumors, however whether considered rare or not the trend would be statistically significant (unadjusted p < 0.001 and Sidak adjusted p = 0.007). Note that in female rats the trends in Pituitary Adenoma and Mammary Gland Carcinoma were also both statistically significant (unadjusted p-values < 0.001 and 0.004, respectively) and remained statistically significant when adjusting for multiplicity (Sidak adjusted p-values 0.001 and 0.017, respectively). However both trends were decreasing in dose. The trend in Parafollicular Cell Adenoma was also barely statistically significant (p = 0.033) and was not statistically significant when adjusted for multiplicity.

3.2.2.2 FDA Reviewer's Results

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

Survival analysis:

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

These results are summarized in the following tables (Tables 8 and 9). The data are grouped for the specified time period, and give the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage is the percent surviving at the end of the interval, as estimated using a Kaplan-Meier estimate on the ungrouped data.

Table 8. Summary of Male Rat Mortality (dose/kg/day)

Period (Weeks)	Control 1	Control 2	Low 50 mg	Medium 150 mg	High 500 mg
1-50	0/50 [1] 100%[2]	1/50 98%	0/50 100%	0/50 100%	3/50 94%
51-78	5/50 90%	2/49 94%	8/50 84%	8/50 84%	11/47 72%
79-91	2/45 86%	3/47 88%	4/42 76%	3/42 78 %	6/36 60%
92-104	4/43 78%	5/44 78%	6/38 64%	6/39 66 %	6/30 48%
Terminal	39	39	32	33	24

[1] number of deaths / number at risk

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

Note that for male rats both the logrank and the Wilcoxon tests of homogeneity in survival were statistically significant (p = 0.0012 and p = 0.0009, respectively), strong evidence of differences in mortality among the treatment groups.

The similar table for females is given below in Table 9:

Table 9. Summary of Female Rat Mortality (dose/kg/day)

Period (Weeks)	Control 1	Control 2	Low 50 mg	Medium 150 mg	High 500 mg
1-50	0/50 [1] 100%[2]	1/50 98%	0/50 100%	4/50 92%	7/50 86%
51-78	6/50 88%	4/49 90%	6/50 88%	9/46 74%	4/43 88%
79-91	7/44 74%	3/45 84%	4/44 80%	2/37 70%	3/39 72%
92-104	9/37 56%	10/42 64%	7/40 66%	6/35 58%	4/36 64%
Terminal	28	32	33	29	32

[1] number of deaths / number at risk

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

For female rats, neither the logrank test of homogeneity in survival curves or the Wilcoxon test were statistically significant (p = 0.7442 and p = 0.5624, respectively).

As with the Sponsor's analyses, there were clear dose related trends in survival in male rats, but no such simple trends in female rats. Strictly speaking, lack of evidence of heterogeneity in survival should not be treated as proof of homogeneity in survival. Nonetheless, it does seem indicative of homogeneity in survival in females.

Tumorigenicity analysis:

Using the Haseman-Lin rules, for an apriori roughly 10% overall error rate, the FDA analysis found statistically significant evidence of tumorigenicity tumorigenicity in liver tumors in both genders and adrenal cortex tumors in males. Using the incidence in the pooled controls, except for carcinoma in females, the liver tumors would be classified as common tumors and the remaining as rare tumors. Then by the Haseman-Lin-Rahman rules described in Section 1.3.1, Statistical Issues, above, in female rats both the trend tests and the pairwise comparisons between the high dose and the pooled controls were statistically significant for both carcinoma and pooled adenoma/carcinoma of the liver (all observed, and unadjusted for multiplicity, $p \leq 0.0041$). Following these same rules, in male rats the trend tests were statistically significant for adenoma, carcinoma, and pooled adenoma/carcinoma in the liver (all $p \leq 0.0044$). The pairwise comparisons between the high dose group and the pooled controls in male rats were statistically significant for carcinoma and pooled adenoma/carcinoma of the liver (both $p \leq 0.0050$). For

adenomas in male rats the pairwise comparison between the high dose group and the pooled controls was close to statistical significance (p = 0.0120). Also, with male rats the trend test in the adrenal cortex was statistically significant (p = 0.0215), and close to significance for the pooled adenomas/carcinomas (p = 0.0269). However the corresponding pairwise tests between the high dose group and the pooled controls in the adrenal cortex were not statistically significant.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5 SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

5.1.1 Statistical Issues

Please see Section 1.3 above.

5.1.2 Collective Evidence

For all statistical analyses, the control groups were pooled. Until near the end of the study, in the male mice the high dose group tended to have the highest mortality, but with no particular dose related trend in the other dose groups and controls (please see Appendix 1 for details). However, as shown in the table below, the omnibus test of treatment differences was not statistically significant (p = 0.7095). In female mice and female rats, there seem to be no consistent differences in mortality among the dose groups, and no other particular dose related pattern in survival. However there were clear differences in survival in male rats. The high dose group consistently had the highest mortality. The low and medium dose groups tracked quite closely, with consistently higher mortality than the pooled controls, but less than the high dose group. For both genders of mice, and also female rats the omnibus tests of homogeneity in survival were not statistically significant. However, differences in survival in male rats were statistically quite significant (p = 0.0012), and correspond to a generally increasing mortality in dose, except that the 50 and 150 mg/kg/day groups are almost coincident.

Table 10. (Identical to table 2) Tests of Homogeneity and Trend in Survival

	Mice			Rats		
Gender	Log Rank	Wilcoxon	Trend	Log Rank	Wilcoxon	Trend
Male	0.7095	0.6318	0.2617	0.0012	0.0009	0.0002
Female	0.5553	0.3735	0.2499	0.7442	0.5624	0.8507

In male and female mice, and in female rats the Bayesian tests of survival seem to be consistent with no particular dose related differences in survival, as is indicated in the classical

frequentist tests above. However, unlike the classical frequentist tests above, in male rats the Bayesian analysis of the model with no treatment group differences seemed to fit better than the models with trend in dose or differences in treatment groups, indicating no difference in survival. However there was weak evidence that the model with a high dose effect and otherwise homogeneous effects fit slightly better than the other models (please see Appendix 2 for details).

For tumorigenicity, in both mouse genders, even without adjusting for multiplicity, there were no statistically significant trends and differences between the pooled controls and the high dose groups. In rats the pattern of statistical significance was more complicated. There was statistically significant evidence of tumorigenicity in liver tumors in both genders and adrenal cortex tumors in males. Using the incidence in the pooled controls, except for carcinoma in females, the liver tumors would be classified as common tumors and the remaining as rare tumors. To control overall Type I error to about 10%, current FDA practice is based on the Haseman-Lin-Rahman rules. Namely, for pairwise tests comparing control to the high dose group, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. For tests of trend rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level.

Then by the Haseman-Lin-Rahman rules, in female rats both the trend tests and the pairwise comparisons between the high dose and the pooled controls were statistically significant for both carcinoma and pooled adenoma/carcinoma of the liver. In male rats the trend tests were statistically significant for adenoma, carcinoma, and pooled adenoma/carcinoma in the liver. The pairwise comparisons between the high dose group and the pooled controls were statistically significant for carcinoma and pooled adenoma/carcinoma of the liver in male rats. For adenomas in male rats the pairwise comparison between the high dose group and the pooled controls was close to statistical significance. With male rats the trend test in the adrenal cortex was statistically significant, and close to significance for the trend test of pooled adenomas/carcinomas.

The following table summarizes incidence and unadjusted p-values of the statistically significant trends and comparisons between the high and pooled control groups. Those comparisons that are statistically significant using the Haseman-Lin-Rahman rules are denoted by an asterix ("*"). Those comparisons close to statistical significance are denoted by a question mark ("?").

Table 11. (Identical to table 3) Tests of Trend and Difference Between the High and Control Groups in Tumorigenicity

Male Mice: None
Female Mice: None

	Pooled Controls	Low	Medium	High	p-values Trend	Hi vs Lo
Male Rats:						
Adrenal/Cortex						
Adenoma	0	0	1	2	0.0215*	0.0941
Adenoma/Carcinoma	0	0	2	2	0.0269?	0.0941
Liver						
Adenoma	2	0	1	5	0.0015*	0.0120?
Carcinoma	4	6	1	7	0.0044*	0.0050*
Adenoma/Carcinoma	6	6	2	12	0.0000*	0.0000*
Female Rats:						
Liver						
Carcinoma	1	1	1	6	0.0008*	0.0041*
Adenoma/Carcinoma	2	3	4	9	0.0001*	0.0004*

*-Statistically Significant (approximately 10-15%)
?-Close to Statistical Significance
Otherwise not statistically significant

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

5.2. Conclusions and Recommendations

In the mouse study there were six treatment groups, i.e., two nominally identically treated control groups, and four dosages of formulation MK-0431 - 50, 125, 250, and 500 mg/kg/day. In the rat study, there were five treatment groups, i.e., two putatively identical control groups, and three dosages of formulation MK-0431 - 50, 150, and 500 mg/kg/day. For analyses, the control groups were pooled. Until near the end of the study the high dose group in male mice could be seen to have the highest mortality rate, but there was no particular dose related trend in the other dose groups and controls (please see Appendix 1 for details). However the tests of homogeneity in survival were not statistically significant, consistent with the hypothesis of homogeneity in survival (Males: Logrank p = 0.7095 & Wilcoxon p = 0.6318). Treatment group related differences in mortality were even less apparent in female mice and rats (all $p \geq 0.3735$). In male rats there was a clear dose related trend in mortality, except that the survival curves in the low and middle dose groups were almost coincident (Male rats: Logrank p = 0.0012 & Wilcoxon p = 0.0009). The Bayesian assessments of mortality seem to be equivocal. Like the usual frequentist tests, the deviance information criteria (DIC) for the assessments in mice and in female rats suggest that there was no difference between treatment groups in mortality. However for male rats the model with no treatment group differences had lower DIC

than the model with trend in dose or differences in treatment groups. Note that the model with a high dose group and homogeneous effects otherwise had a slightly lower DIC, and the posterior distribution of the parameter corresponding to a high dose group was clearly bounded away from zero (please see Appendix 2 for details).

For tumorigenicity, in both mouse genders, even without adjusting for multiplicity, there were no statistically significant trends and differences between the pooled controls and the high dose groups. In rats the pattern of statistical significance was more complicated. There was statistically significant evidence of tumorigenicity in liver tumors in both genders and adrenal cortex tumors in males. Using the incidence in the pooled controls, except for carcinoma in females, the liver tumors would be classified as common tumors and the remaining as rare tumors. Then by the Haseman-Lin-Rahman rules described in Section 1.3.1, Statistical Issues, above, in female rats both the trend tests and the pairwise comparisons between the high dose and the pooled controls were statistically significant for both carcinoma and pooled adenoma/carcinoma of the liver (all observed, and unadjusted for multiplicity, $p \leq 0.0041$). Following these same rules, in male rats the trend tests were statistically significant for adenoma, carcinoma, and pooled adenoma/carcinoma in the liver (all $p \leq 0.0044$). The pairwise comparisons between the high dose group and the pooled controls were statistically significant for carcinoma and pooled adenoma/carcinoma of the liver in male rats (both $p \leq 0.0050$). For adenomas in male rats the pairwise comparison between the high dose group and the pooled controls was close to statistically significance ($p = 0.0120$). With male rats the trend tests in the adrenal cortex was statistically significant ($p = 0.0215$), and close to significance for the pooled adenomas/carcinomas ($p = 0.0269$). However the corresponding pairwise tests between the high dose group and the pooled controls in the adrenal cortex were not statistically significant.

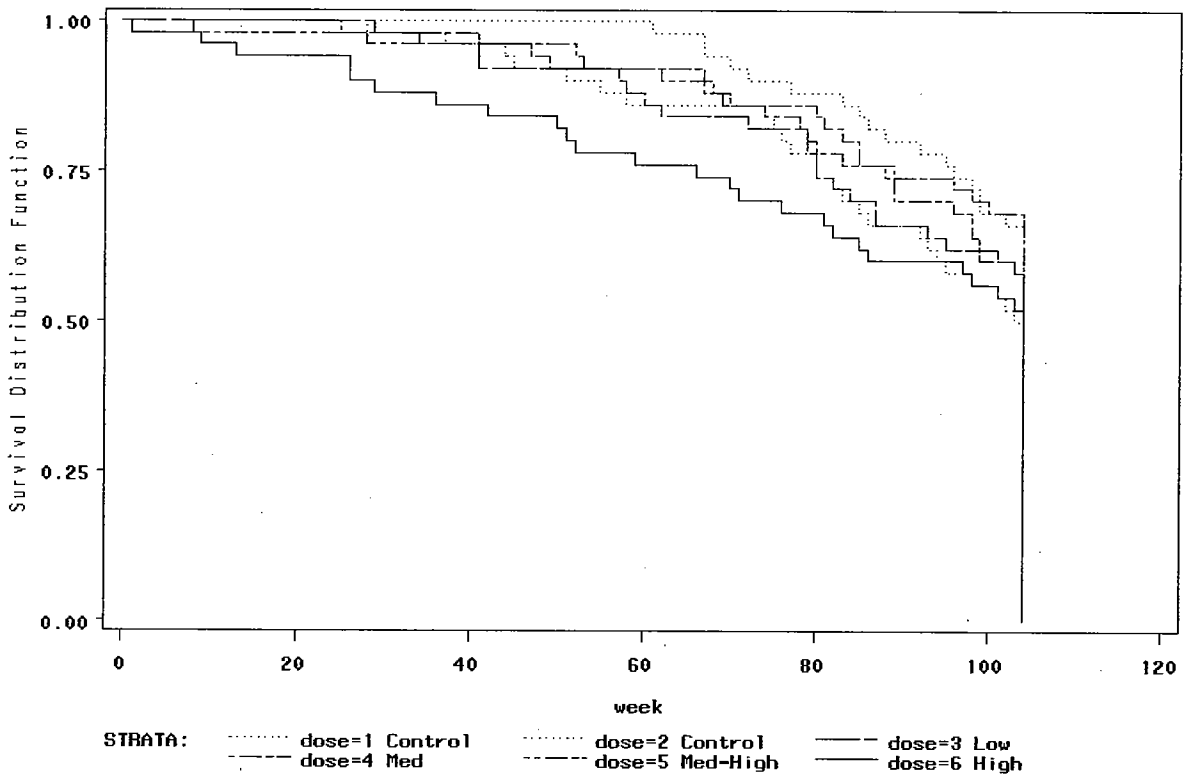
APPENDICES:

Appendix 1. Survival Analysis

The omnibus tests of heterogeneity in survival among the treatment groups, including the pooled controls, were only statistically significant in male rats (Mice Males: Logrank p = 0.7095 & Wilcoxon p = 0.6318, Mice Females: Logrank p = 0.5553 & Wilcoxon p = 0.3735, Rat Males: Logrank p = 0.0012 & Wilcoxon p = 0.0009, and Rat Females: Logrank p = 0.7442 & Wilcoxon p = 0.5624). Note the Cox model provides a test of trend that generalizes the logrank test for comparing survival in two treatments (Mice Males: p = 0.2617, Mice Females: p = 0.2499, Rat Males: p = 0.0002, and Rat Females: p = 0.8507). Again the test of trend in male rats was highly statistically significant.

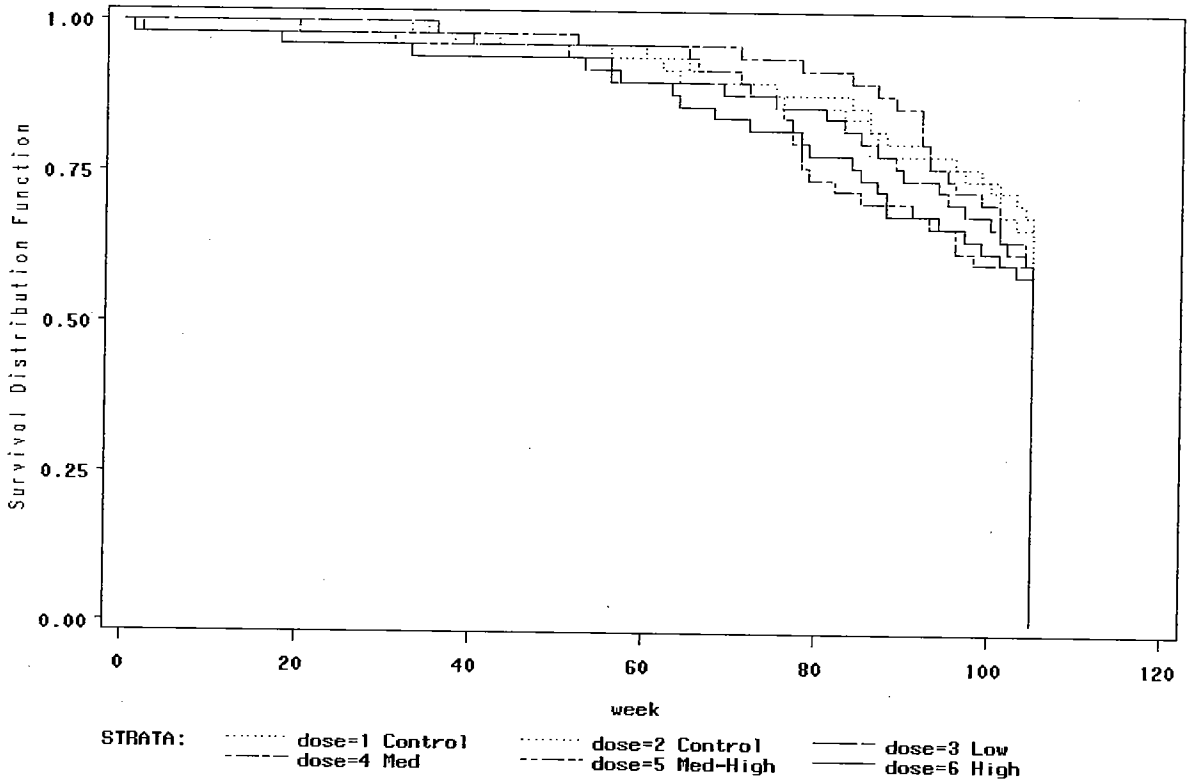
The figures below display the Kaplan-Meier estimated survival curves for the four different species by gender combinations.

Figure A.1.1 Male Mice



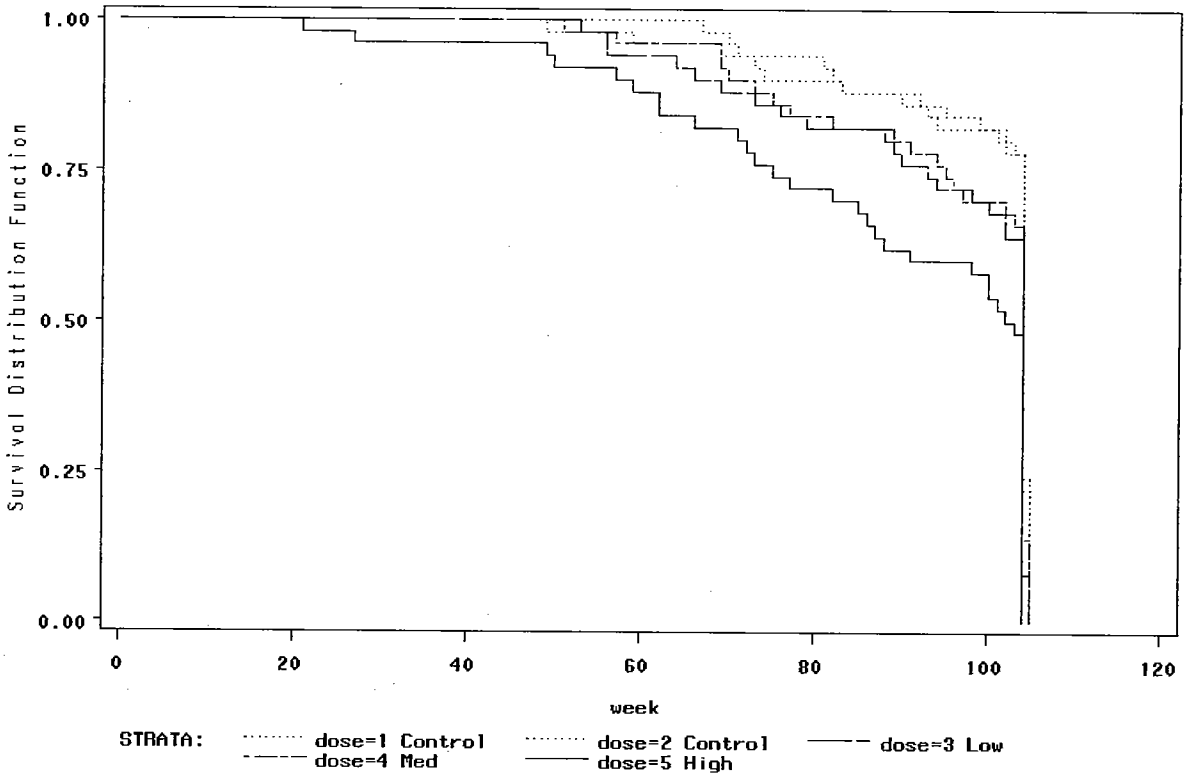
Thus, in these plots the control doses are denoted by dotted lines while the high dose group is solid.

Figure A.1.2 Female Mice



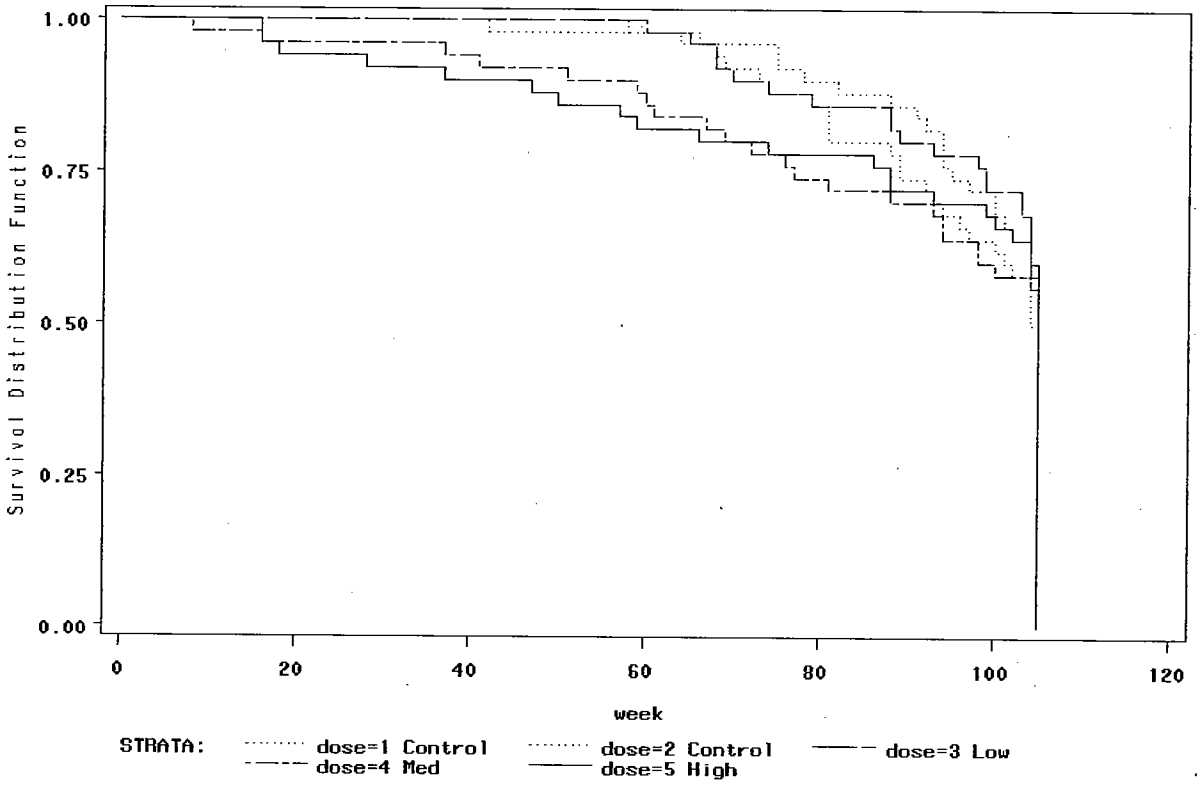
Note that for neither gender was the test in survival as a function of dose statistically significant ($p = 0.2617$ for males and $p = 0.2499$ for females).

Figure A.1.3 Male Rats



As noted above the test of homogeneity in survival in males was statistically significant (Logrank p = 0.0012, Wilcoxon p = 0.0009, and Cox: p = 0.0002).

Figure A.1.4 Female Rats



For females note that survival as a function of dose was never statistically significant (Logrank p = 0.7442, Wilcoxon p=0.5624, and Cox test of trend: p = 0.8507).

Appendix 2. Bayesian Analysis of Survival

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

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

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

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

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

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

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

Note that without other information we would expect the treatment effects in the control groups to be exchangeable (i.e., effectively the treatment groups can be treated as identical). Then, after pooling the two control groups, in the mouse study there were four treatment groups and in the rat study there were five.

Frequentist analysis of this model uses asymptotics to analyze the linear predictor, ignoring the baseline hazard $h_0(t)$. A Bayesian analysis requires priors on all parameters, including the baseline hazard. Perhaps the simplest Bayesian model would postulate a within interval constant baseline hazard. That is, suppose the time axis can be partitioned as $(a_1=0, a_2]$, $(a_2, a_3]$, . . ,$(a_T,a_{T+1}]$. Assume the constant baseline hazard λ_j for observations in $(a_j,a_{j+1}]$. For simplicity we assume the intervals are of equal length, with a Gamma prior, i.e.,

$\lambda_j \sim$ Gamma (α_j,γ_j),

where for this analysis we assume that the gamma distribution parameters in the prior are equal for each interval (i.e., $\alpha_1 = \alpha_2 = \ldots = \alpha_T$ and $\gamma_1 = \gamma_2 = \ldots = \gamma_T$.). Since the last time period involves a terminal sacrifice under control of the experimenter, one might argue that this assumption of equal parameters at all intervals in the prior may not be appropriate, but it will have little difference on the results and is a convenient way to model ignorance. For this analysis the gamma prior chosen had a mean of 1 event with a variance of 100.

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

Using so called dummy coding, we can define, for each treatment group k,
$\delta_k = 1$ for the ith treatment group,
0 otherwise.

Then three possibly relevant models for treatment effect could be expressed as follows:
(1) Parameterization of a different effect for each treatment (with 5 treatments),
$x_i^t\beta = \beta_0 + \beta_1 * \delta_1 + \beta_2 * \delta_2 + \beta_3 * \delta_3 + \beta_4 * \delta_4$.
(2) Parameterization of a linear effect of measures dose over treatment groups,
$x_i^t\beta = \beta_0 + \beta_1 *$ dose
(3) Parameterization of no differences in survival across treatment groups, $x_i^t\beta = \beta_0$.

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

Finally, although this is a post hoc model specification based on the observed Kaplan-Meier curves, we have the model for male rats, which differentiates the high dose group from the others:
4) $x_i^t\beta = \beta_0 + \beta_4 * \delta_4$.

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

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

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

Then the likelihood for subject i can be written as:

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

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

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

Note that for intervals above a_j, $-e^{x'\beta}\gamma_{ik} = 0$, so $\exp(-e^{x'\beta}\gamma_{ik})$ does not contribute to the product.

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

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

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

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

Deviance Information Criterion for Mice	Males	Females
Model with heterogeneity over the five treatment groups	29.387	27.282
Model with linear trend in dose groups, 0=vehicle.	26.360	24.199
Model with constant dose effect	25.237	23.093

Deviance Information Criterion for Rats	Males	Females
Model with heterogeneity over the four treatment groups	19.733	24.112
Model with linear trend in dose groups, 0=vehicle.	17.751	22.046
Model with constant dose effect	16.699	21.116
Model with high dose effect ≠ medium = low = vehicle	16.674	

Using the DIC, for females of both species, and for male mice, the models with no treatment effect on survival seems to fit best. For male rats the model with a high dose effect different from the other effects fits slightly better than the model with no treatment effects, though such a small difference is often ignored. The tables below summarize the estimated posterior distributions of the treatment parameters. For male mice and for females in both species, zero is located well within the approximate credible intervals (i.e., with lower endpoint in the 2.5% column and upper endpoint in the 97.5% column) for each parameter. This indicates

that for these animals there was no evidence of differences from the pooled controls or in overall trend in survival among the treatment groups. In male rats there was some evidence of treatment related differences in survival (please see Table A.2.2). These results are displayed in more detail in the following summaries of the posterior distributions of the parameters:

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

Male testing homogeneity over five parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.02947	0.2682	0.00344	-0.5077	0.03117	0.5474	4001	16000
beta[2]	-0.2667	0.2898	0.003573	-0.8505	-0.2617	0.2859	4001	16000
beta[3]	-0.0125	0.2699	0.003511	-0.5557	-0.006952	0.5085	4001	16000
beta[4]	0.2771	0.2609	0.003496	-0.2369	0.2781	0.7848	4001	16000

Male model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	2.785E-4	2.388E-4	2.856E-6	-1.994E-4	2.794E-4	7.342E-4	4001	15997

Female testing homogeneity over five parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.2666	0.2853	0.003894	-0.2987	0.2693	0.8126	4001	16000
beta[2]	0.1562	0.2857	0.003785	-0.4138	0.1602	0.7028	4001	16000
beta[3]	0.2701	0.2865	0.004037	-0.3088	0.2713	0.8223	4001	16000
beta[4]	0.3233	0.2834	0.003798	-0.2355	0.327	0.8726	4001	16000

Female model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	2.369E-4	2.459E-4	3.026E-6	-2.589E-4	2.411E-4	7.093E-4	4001	15998

For rats the posterior summaries of the parameter values are given in Table A.2.2 below. In male rats, the parameter diff = $\beta_3 - \beta_2$, assesses difference in the linear predictor between the medium dose group and the low dose group, and suggests there is no difference between these groups. However, in male rats both the posterior distribution of the trend and the difference between the high dose groups place most of the probability away from zero, indicating high probability that these are not zero

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

Male testing homogeneity over four parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.5732	0.3225	0.004996	-0.06759	0.5773	1.211	4001	16000
beta[2]	0.4919	0.3285	0.00524	-0.1627	0.494	1.129	4001	16000
beta[3]	1.102	0.2923	0.005032	0.5318	1.102	1.682	4001	16000
diff	-0.08128	0.343	0.002559	-0.7544	-0.0812	0.5963	4001	16000

Male model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	0.0186	0.005189	6.032E-5	0.008221	0.01869	0.02857	4001	16000

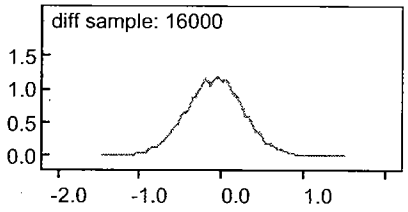
Female testing homogeneity over four parameter groups

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	-0.1951	0.2937	0.004003	-0.7922	-0.1885	0.3646	4001	16000
beta[2]	0.1706	0.274	0.003726	-0.3779	0.1756	0.7006	4001	16000
beta[3]	-0.02894	0.2885	0.003809	-0.6081	-0.02334	0.5167	4001	16000

Female model for simple trend in dose

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	3.791E-4	0.0055	6.636E-5	-0.01066	5.147E-4	0.01061	4001	16000

The estimated posterior density of the diff parameter in males is given below:



Again, this indicates that in this model there is no strong evidence of differences between the low dose group, and the medium dose group. However, in male rats the approximate 95% credible interval for the difference between the high dose group and the pooled controls is [0.53, 1.68], quite different from 0.0. The approximate 95% credible interval for the linear trend parameter is [0.008, 0.286], also bounded away from 0. Both conclusions are consistent with an increasing mortality in the high dose group.

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

2 Page(s) Withheld

✓ § 552(b)(4) Trade Secret / Confidential

_____ § 552(b)(5) Deliberative Process

_____ § 552(b)(5) Draft Labeling

Withheld Track Number: Statistical-2

Appendix 3. Sponsor's Tumorigenicity Analyses

For each gender within each species, for each neoplasm within each organ, the Sponsor provided tables of tumor incidence and the results of Peto tests of trend over the dose groups and the pooled controls. As noted before, for testing trend, controls were pooled. For the Haseman-Lin-Rahman rules the incidence in the pooled controls will be used to determine if a tumor is rare or common. For statistically significant outcomes the Sponsor provides both the unadjusted p-value and p-value adjusted for multiplicity using Sidak's inequality. Note, however, that the current Agency analyses follow the Haseman-Lin-Rahman adjustment for multiplicity (please see Section 3.1 or Appendix 4 for discussion).

The only statistically significant trends in mice were negative in dose (i.e., decreasing trend in tumorigencity with increasing dose), and do not remain statistically significant after adjusting for multiplicity (see adrenal cortex adenoma in males and primary site undetermined leukemia in females in Tables A.3.1 and A.3.2). In male rats there were statistically significant trends in Liver Hepatocellular Adenomas and Carcinomas (unadjusted $p = 0.002$ and $p = 0.006$, respectively). Note these would be classed as common tumors. Adjusting for multiplicity using the Haseman-Lin-Rahman rules or the Sponsor's Sidak adjustment only the trend in Liver Hepatocellular Adenomas remained statistically significant (Sidak adjustment $p = 0.016$). Using either adjustment Liver Hepatocellular Carcinomas in male rats remained statistically significant (Sidak adjustment $p = 0.066$). In males the trend in Islet cell adenoma was (unadjusted) statistically significant ($p = 0.023$), but negative in trend in dose, while adenomas in the adrenal cortex were positive in trend, but again, barely statistically significant ($p = 0.023$). Using either the Haseman-Lin-Rahman rules or the Sidak correction to adjust for the multiplicity in tests, neither trend remained statistically significant. In female rats Hepatocellular Carcinomas would be classed as rare tumors, however whether considered rare or not the trend would be statistically significant (unadjusted $p < 0.001$ and Sidak adjusted $p = 0.007$). Note that in female rats the trends in Pituitary Adenoma and Mammary Gland Carcinoma were also both statistically significant (unadjusted p-values <0.001 and 0.004, respectively) and remained statistically significant when adjusting for multiplicity (Sidak adjusted p-values 0.001 and 0.017, respectively). However both trends were decreasing in dose. The trend in Parafollicular Cell Adenoma was also barely statistically significant ($p = 0.033$) and was not statistically significant when adjusted for multiplicity.

Tables A.3.1 and A.3.2 summarize tumor incidence in mice for treatment group. The results of an overall test of trend are displayed in the last column. Note that p-values that are statistically significant at the conventional 0.05 level are given in bold, while with negative trends the p-value is flagged with an "N".

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

Table A.3.1. TT #03-615-0,-2 Tumor Incidence and Tests of Incidence in Male Mice

Males	Control 1	Control 2	50 MKD	125 MKD	250 MKD	500 MKD	Trend P-value
Adrenal							
Benign Spindle Cell Tumor	4	1	2	4	1	-	0.054N
Malignant Spindle Cell Tumor	-	-	1	1	-	-	0.434N
Cortex, Adenoma	3	4	5	5	4	-	**0.049N** (0.573N)
Skin							
Fibroma	-	-	-	-	1	1	0.071
Fibrosarcoma	1	1	2	-	2	-	0.290N
Hemangiosarcoma	-	-	2	-	1	-	0.440N
Malignant Histiocytoma Fibrous	-	-	1	-	1	-	0.422
Rhabdomyosarcoma	-	-	2	-	1	-	0.443N
Malignant Scwannoma	1	1	-	-	1	-	0.362N
Primary Site Undetermined							
Histiocytic Sarcoma	1	1	1	-	-	-	0.099N
Leukemia	-	2	-	-	-	-	0.115N
Malignant Mastocytoma	1	1	-	1	-	-	0.187N
Lymphoma	4	3	3	2	1	1	0.079N
Liver							
Hemangiosarcoma	1	1	1	2	1	-	0.248N
Hepatocellular Adenoma	11	2	8	8	11	9	0.088
Hepatocellular Carcinoma	8	4	2	4	2	4	0.374N
Thymus							
Benign Thymoma	-	1	2	-	-	-	0.122N
Lung							
Adenocarcinoma	10	8	4	7	7	4	0.127N
Adenoma	10	7	14	20	15	8	0.455
Kidney							
Tubule, Adenoma	-	-	2	3	-	2	0.172
Pituitary							
Anterior Lobe, Adenoma	-	-	2	-	1	1	0.172
Testis							
Hemangioma	-	1	-	1	-	-	0.412N
Leydig Cell, Adenom	2	2	1	1	-	1	0.242N
Rete Testis, Adenoma	-	1	-	-	-	1	0.303
Epididymus, Benign Interstitial Cell Tumor	-	1	-	-	-	1	0.303
Epididymus, Malignant Interstitial Cell Tumor	1	-	-	2	-	-	0.351N
Eye- Harderian Gland							
Adenocarcinoma	-	1	-	-	-	1	0.254
Adenoma	6	1	1	9	2	3	0.488
Pancreas							
Islet, Adenoma	-	-	1	-	1	-	0.421
Thyroid							
Follicular Cell, Adenoma	-	-	1	1	-	-	0.438N

Table A.3.2. TT #03-615-0,-2 Tumor Incidence and Tests of Incidence in Female Mice

Females	Control 1	Control 2	50 MKD	125 MKD	250 MKD	500 MKD	Trend P-value
Primary Site Undetermined							
Malignant Plasma Cell Tumor	-	-	1	-	1	1	0.123
Histiocytic Sarcoma	-	1	4	1	-	2	0.386
Leukemia	1	2	1	-	-	-	**0.041N** (0.435N)
Lymphoma	8	17	7	15	11	11	0.372
Liver							
Hepatocellular Adenoma	1	2	1	2	2	3	0.114
Hepatocellular Carcinoma	1	-	-	-	1	-	>0.500
Skin							
Fibrosarcoma	-	-	2	1	2	-	0.498
Leiomyosarcoma	-	-	-	1	-	1	0.123
Liposarcoma	-	-	-	1	1	-	0.361
Rhabdomyosarcoma	1	1	-	-		-	0.128N
Pituitary							
Anterior Lobe, Adenoma	4	3	2	7	2	1	0.141N
Uterus							
Adenomacarcinoma	-	1	-	1	-	-	0.390N
Leiomyoma	-	3	-	2	1	-	0.216N
Leiomyosarcoma	-	1	1	1	1	-	0.388N
Polyp	2	4	3	3	3	1	0.153N
Stromal Sarcoma	-	1	-	-	2	-	0.468
Malignant Schwannoma	-	1	1	-	1	-	0.428N
Ovary							
Cystadenoma	3	1	2	2	2	3	0.248
Benign Luteoma	2	3	6	4	-	2	0.163N
Malignant Theca Cell Tumor	-	-	-	1	-	1	0.170
Lung							
Adenocarcinoma	4	6	3	2	4	2	0.195N
Adenoma	7	8	7	6	4	5	0.174N
Eye- Harderian Gland							
Adenoma	2	5	3	-	4	4	0.305
Adrenal							
Malignant Pheochromocytoma	-	1	-	-	-	1	0.310
BenignSpindle Cell Tumor	-	1	-	1	-	-	0.398N
Mammary Gland							
Malignant Pheochromocytoma	-	-	1	1	1	-	>0.500
BenignSpindle Cell Tumor	1	3	2	3	3	1	0.398N
Small Intestine							
Adenocarcinoma	1	-	-	-	1	-	>0.500
Spleen							
Hemangiosarcoma	-	1	-	-	1	-	>0.500

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

Table A.3.3 TT #03-097-0 Tumor Incidence and Tests of Incidence in Male Rats

Males	Control 1 1	Control 1 2	50 MKD	150 MKD	500 MKD	Trend P-value
Liver						
Hepatocellular Adenoma	1	1	-	1	5	**0.002 (0.016)**
Hepatocellular Carcinoma	1	3	6	1	7	**0.006** (0.066)
Pancreas						
Islet, Adenoma	-	4	3	-	-	**0.023N** (0.141N)
Islet, Carcinoma	1	2	2	1	-	0.107N
Adrenal						
Benign Pheochromocytoma	3	5	3	5	3	0.489N
MalignantPheochromocytoma	1	1	-	-	-	0.219N
Cortex Adenoma	-	-	-	1	2	**0.023** (0.293)
Thyroid						
Parafollicular Cell, Adenoma	3	6	5	5	2	0.213N
Parafollicular Cell, Carcinoma	-	-	2	2	1	0.228
Follicular Cell, Adenoma	-	-	1	-	1	0.132
Follicular Cell, Carcinoma	1	1	1	-	2	0.167
Mammary Gland						
Carcinona	-	-	1	-	1	0.142
Testis						
Benign Interstitial Cell Tumor	3	3	3	-	1	0.149N
Primary Site Undetermined						
Leukemia	-	-	1	-	1	0.150
Lymphoma	1	-	-	-	1	0.294
Skin						
Benign Basal Cell Tumor	1	1	-	-	1	0.389
Malignant Basal Cell Tumor	1	-	-	-	1	0.314
Fibroma	1	-	1	-	-	0.378N
Benign Keratocanthoma	1	1	1	1	-	0.332N
Squamous Papilloma	1	-	1	1	-	0.464N
Sarcoma	1	1	1	-	-	0.194N
Parathyroid						
Adenoma	1	-	-	1	1	0.240
Pituitary						
Adenoma	27	20	21	25	19	0.331
Brain						
Astrocytoma	-	-	1	1	-	0.379

Table A.3.4 TT #03-097-0 Tumor Incidence and Tests of Incidence in Female Rats

Females	Control 1	Control 2	50 MKD	150 MKD	500 MKD	Trend P-value
Liver						
Hepatocellular Adenoma	1	-	2	3	3	0.090
Hepatocellular Carcinoma	-	1	1	1	6	**<0.001 (0.007)**
Pituitary						
Adenoma	34	42	35	28	19	**0.001N (0.001N)**
Mammary Gland						
Adenoma	1	-	-	1	1	0.495
Carcinoma	17	16	10	10	5	**0.004N (0.017N)**
Fibroadenoma	8	8	9	6	7	0.473N
Thyroid						
Parafollicular Cell, Adenoma	4	1	6	2	-	0.135
Parafollicular Cell, Carcinoma	2	-	1	2	1	0.470
Follicular Cell, Adenoma	-	-	2	-	-	>0.500
Brain						
Astrocytoma	1	1	1	1	-	0.316N
Benign Granular Cell Tumor	-	-	-	1	1	0.119
Vagina						
Malignant Granular Cell Tumor	-	2	-	-	-	0.153N
Skin						
Squamous Papilloma	1	1	-	-	-	0.180N
Adrenal						
Benign Pheochromocytoma	1	1	2	-	1	0.473
Cortex Adenoma	-	2	-	-	-	0.208N
Cortex Carcinoma	-	-	2	-	-	0.388N
Primary Site Undetermined						
Histiocytic Sarcoma	-	1	-	1	1	0.249
Uterus						
Endometrial Stromal Polyp	1	2	4	1	1	0.408N
Cervix, Endometrial Stromal Polyp	1	2	-	1	-	0.255N
Cervix, Endometrial Stromal Sarcoma	-	1	-	1	-	>0.500
Heart						
Endocardium, Malignant Schwannoma	-	1	1	-	-	0.323N
Ovary						
Benign Sertoli Cell Tumor	1	1	-	1	-	0.334N
Parathyroid						
Adenoma	1	-	-	-	1	0.371
Pancreas						
Islet, Adenoma	-	1	1	-	-	0.397N

Appendix 4. FDA Tumorigenicity Analysis

Tables A.4.1 and A.4.2 below display the number of neoplasms in each organ and tumor combination in mice taken from the datasets provided by the Sponsor. Tables A.4.3 and A.4.4 below display similar results for rats. For each dose group, the tables present the number of animals where histopathological analysis detected a tumor. The Sponsor states that all animals were analyzed, so in each treatment group, all 50 animals were assessed for each tumor.

The p-values correspond to tests of trend among the pooled controls and the treatment groups and a comparison of the high dose group with the pooled controls. For 10 or fewer animals being tested, the reported significance levels are from exact tests (i.e., assuming the marginal totals for the number of animals with and without the neoplasm are fixed). Otherwise the significance levels are from asymptotic tests.

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

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

So, for example, for a rare tumor (with incidence in the pooled control groups ≤ 1%, i.e., 0 or 1 tumor), a trend would be considered statistically significant if the computed significance level was at or less than 0.025, while a comparison between the high dose group and the pooled controls (i.e., a pairwise comparison) would be statistically significant if the computed significance level was no more than 0.05.

In mice, in both genders, even without adjusting for multiplicity, there were no statistically significant trends and differences between the pooled controls and the high dose groups. In rats, as also noted in Appendix 3, the pattern of statistical significance was more complicated. There was statistically significant evidence of tumorigenicity in liver tumors in both genders and adrenal cortex tumors in males. Using the incidence in the pooled controls, except for carcinoma in females, the liver tumors would be classified as common tumors and the remaining as rare tumors. Then by the Haseman-Lin-Rahman rules, in female rats both the trend tests and the pairwise comparisons between the high dose and the pooled controls were statistically significant for both carcinoma and pooled adenoma/carcinoma of the liver (all $p \leq 0.0041$). Following these same rules, in male rats the trend tests were statistically significant for adenoma, carcinoma, and pooled adenoma/carcinoma in the liver (all $p \leq 0.0044$). The pairwise comparisons between the high dose group and the pooled controls were statistically significant for carcinoma and pooled adenoma/carcinoma of the liver in male rats (both $p \leq 0.0050$). For

adenomas in male rats the pairwise comparison between the high dose group and the pooled controls was close to statistical significance (p = 0.0120). With male rats the trend tests in the adrenal cortex was statistically significant (p = 0.0215), and close to significance for the pooled adenomas/carcinomas (p = 0.0269). However the corresponding pairwise tests between the high dose group and the pooled controls in the adrenal cortex were not statistically significant.

The following tables show the tumor incidence and the significance levels of the tests of trend and also of the high dose group versus the pooled controls. When there are no observed values in the controls and the high dose group, the test of differences is not defined and thus no p-value is given.

Table A.4.1. Tumorgenicity in Male Mice

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	Med-High	High	p-values: Trend	Hi vs Cntl
Adrenal								
Spindle Cell Tumor	4	1	3	5	1	0	0.9538	1.0000
Adrenal/Cortex								
Adenoma	3	4	5	5	4	0	0.9513	1.0000
Bone Marrow								
Hemangiosarcoma	0	0	1	0	0	0	0.6705	
Coagulating Gland								
Adenoma	0	0	0	0	1	0	0.3125	
Eye/Harderian Gland								
Adenocarcinoma	0	1	0	0	0	1	0.2513	0.5055
Adenoma	6	1	1	9	2	3	0.5001	0.5977
Kidney								
Hemangiosarcoma	0	0	0	0	0	1	0.1459	0.2957
Kidney/Tubule								
Adenoma	0	0	2	3	0	2	0.1710	0.0932
Liver								
Hemangioma	0	0	0	0	1	0	0.3125	
Hemangiosarcoma	1	1	1	2	1	0	0.8055	1.0000
Hemangioma/-sarcoma	1	1	1	2	2	0	0.7094	1.0000
Hepatoblastoma	1	0	0	0	0	0	1.0000	1.0000
Hepatocellular adenoma	11	2	8	8	11	9	0.0896	0.2102
Hepatocellular carcinoma	8	4	2	4	2	4	0.6387	0.7218
Liver/Bile Duct								
Cystadenoma	0	0	0	1	0	0	0.4063	
Liver/Ito Cell								
Adenoma	0	1	0	0	0	0	1.0000	1.0000
Lung								
Adenocarcinoma	10	8	4	7	7	4	0.8764	0.9621
Adenoma	10	7	14	20	15	8	0.4622	0.4724
Lymph Node								
Hemangiosarcoma	0	0	0	1	0	0	0.4063	
Pancreas/Islet								
Adenoma	0	0	1	0	1	0	0.4194	
Parathyroid								
Adenoma	0	0	0	0	1	0	0.3125	

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

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	Med-High	High	p-values: Trend	Hi vs Cntl
Pituitary/Anterior Lobe								
Adenoma	0	0	2	0	1	1	0.2245	0.3095
Pituitary/Pars Intermedia								
Carcinoma	0	0	0	0	1	0	0.3120	
Primary Site Undetermined								
Histiocytic Sarcoma	1	1	1	0	0	0	0.9597	1.0000
Leukemia	0	2	0	0	0	0	1.0000	1.0000
Lymphoma	4	3	3	2	1	1	0.9167	0.9438
Mastocytoma	1	1	0	1	0	0	0.8754	1.0000
Osteosarcoma	1	0	0	0	0	0	1.0000	1.0000
Skeletal Muscle								
Hemangiosarcoma	0	0	0	0	0	1	0.1477	0.3095
Skin								
Carcinoma	0	0	1	0	0	0	0.6558	
Fibroma	0	0	0	0	1	1	0.0875	0.2105
Fibrosarcoma	1	1	2	0	2	0	0.7409	1.0000
Hemangiosarcoma	0	0	2	0	1	0	0.5064	
Histiocytoma	0	0	1	0	1	0	0.4194	
Lipoma	1	0	0	0	0	0	1.0000	1.0000
Rhabdomyosarcoma	0	0	2	0	1	0	0.5661	
Schwannoma	1	1	0	0	1	0	0.7401	1.0000
Sebaceous adenoma	0	0	1	0	0	0	0.6705	
Small Intestine								
Hemangioma	1	0	0	0	0	0	1.0000	1.0000
Spleen								
Hemangiosarcoma	0	0	0	0	1	0	0.3125	
Stomach/Glandular mucosa								
Adenoma	0	0	0	1	0	0	0.4706	
Stomach/Nonglandular mucosa								
Papilloma	1	0	0	0	0	0	1.0000	1.0000
Systemic								
Hemangioma	1	1	0	1	1	0	0.6961	1.0000
Hemangiosarcoma	2	1	4	3	3	2	0.4016	0.4873
Hemangioma/-sarcoma	3	2	4	4	4	2	0.5150	0.6621
Testis								
Hemangioma	0	1	0	1	0	0	0.7065	1.0000
Testis/Epididymis								
Interstitial Cell Tumor	1	1	0	2	0	1	0.4832	0.6762
Testis/Leydig Cell								
Adenoma	2	2	1	1	0	1	0.7924	0.8516
Testis/Rete Testis								
Adenoma	0	1	0	0	0	1	0.3007	0.5258
Thymus								
Thymoma	0	1	2	0	0	0	0.9042	1.0000
Thyroid/Follicular Cell								
Adenoma	0	0	1	1	0	0	0.6233	
Urinary Bladder								
Hemangiosarcoma	1	0	0	0	0	0	1.0000	1.0000
Urinary Bladder/Transitional E								
Papilloma	1	0	0	0	0	0	1.0000	1.0000

Table A.4.2. Tumorgenicity in Female Mice

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	Med-High	High	p-values: Trend	Hi vs Cntl
Adrenal								
Pheochromocytoma	0	1	0	0	0	1	0.5644	0.5107
Spindle Cell Tumor	0	1	0	1	0	0	0.3754	1.0000
Bone								
Chondroma	0	0	0	1	0	0	0.0833	
Osteoma	0	0	0	0	0	1	0.3211	0.2990
Osteosarcoma	1	0	0	0	0	0	1.0000	1.0000
Brain								
Meningioma	1	0	0	0	0	0	1.0000	1.0000
Ear/Zymbal's Gland								
Carcinoma	0	0	0	0	1	0	0.4789	
Eye/Harderian Gland								
Adenocarcinoma	0	0	0	0	1	0	0.4789	
Adenoma	2	5	3	0	4	4	0.9439	0.5048
Kidney								
Mesenchymal tumor	0	0	0	0	1	0	0.4815	
Kidney/Tubule								
Carcinoma	0	1	0	0	0	0	1.0000	1.0000
Large Intestine								
Leiomyosarcoma	0	0	0	0	1	0	0.4789	
Liver								
Hepatocellular adenoma	1	2	1	2	2	3	0.3478	0.2547
Hepatocellular carcinoma	1	0	0	0	1	0	0.7298	1.0000
Liver/Bile Duct								
Cystadenoma	0	1	0	0	0	0	1.0000	1.0000
Lung								
Adenocarcinoma	4	6	3	2	4	2	0.9063	0.9334
Adenoma	7	8	7	6	4	5	0.7110	0.8367
Lymph Node								
Hemangioma	0	0	0	0	1	0	0.4789	
Hemangiosarcoma	0	0	0	0	1	0	0.4789	
Hemangioma/-sarcoma	0	0	0	0	2	0	0.5644	
Mammary Gland								
Adenoacanthoma	0	0	1	1	1	0	0.2404	
Adenocarcinoma	1	3	2	3	3	1	0.3971	0.8372
Ovary								
Cystadenoma	3	1	2	2	2	3	0.4255	0.3951
Hemangioma	0	0	0	0	0	1	0.2941	0.4375
Hemangiosarcoma	1	0	0	0	0	0	1.0000	1.0000
Hemangioma/-sarcoma	1	0	0	0	0	1	0.5882	0.7000
Leiomyoma	1	0	0	0	0	0	1.0000	1.0000
Luteoma	2	3	6	4	1	2	0.3889	0.6774
Theca cell tumor	0	0	0	1	0	1	0.1141	0.4286
Pancreas/Islet								
Adenoma	1	0	0	0	0	0	1.0000	1.0000
Pituitary/Anterior Lobe								
Adenoma	4	3	2	7	2	1	0.0398	0.9481

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

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	Med-High	High	p-values: Trend	Hi vs Cntl
Primary Site Undetermined								
Hemangiosarcoma	0	0	0	1	0	0	0.1684	
Histiocytic Sarcoma	0	1	4	1	0	2	0.5417	0.2259
Leukemia	1	2	1	0	0	0	0.9859	1.0000
Lymphoma	8	17	7	15	11	11	0.1076	0.5749
Mastocytoma	0	0	0	0	0	1	0.3211	0.2990
Plasma Cell Tumor	0	0	1	0	1	1	0.5240	0.3103
Sarcoma	0	0	0	1	0	0	0.1733	
Schwannoma	0	1	0	0	0	0	1.0000	1.0000
Skin								
Fibrosarcoma	0	0	2	1	2	0	0.4130	
Histiocytoma	0	0	0	0	0	1	0.3333	0.4286
Keratoacanthoma	0	1	0	0	0	0	1.0000	1.0000
Leiomyosarcoma	0	0	0	1	0	1	0.0794	0.3000
Lipoma	0	0	1	0	0	0	0.6421	
Liposarcoma	0	0	0	1	1	0	0.1495	
Myxosarcoma	1	0	0	0	0	0	1.0000	1.0000
Papilloma	0	0	0	0	0	1	0.3211	0.2990
Rhabdomyosarcoma	1	1	0	0	0	0	1.0000	1.0000
Small Intestine								
Adenocarcinoma	1	0	0	0	1	0	0.7259	1.0000
Spleen								
Hemangiosarcoma	0	1	0	0	1	0	0.7606	1.0000
Systemic								
Hemangioma	0	0	0	1	2	1	0.2067	0.4375
Hemangiosarcoma	1	1	0	1	2	0	0.5194	1.0000
Hemangioma/-sarcoma	1	1	0	2	4	1	0.2859	0.7938
Urinary Bladder/Transitional E								
Papilloma	0	0	0	0	0	1	0.3211	0.2990
Uterus								
Adenocarcinoma	0	1	0	1	0	0	0.3092	1.0000
Granular cell tumor	0	0	0	1	0	0	0.1684	
Hemangioma	0	0	0	0	1	0	0.4861	
Leiomyoma	0	3	0	2	1	0	0.3112	1.0000
Leiomyosarcoma	0	1	1	1	1	0	0.3926	1.0000
Polyp	2	4	3	3	3	1	0.5455	0.9365
Sarcoma	0	1	0	0	2	0	0.7386	1.0000
Schwannoma	0	1	1	0	1	0	0.8060	1.0000
White Adipose Tissue								
Hemangioma	0	0	0	1	0	0	0.2973	
Lipoma	0	0	1	0	0	0	0.7027	

Table A.4.3. Tumorgenicity in Male Rats

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	High	p-values: Trend	Hi vs Cntrl
sex=M							
Adrenal							
Pheochromocytoma	4	6	3	5	3	0.6282	0.7536
Adrenal/Cortex							
Adenoma	0	0	0	1	2	0.0215	0.0941
Carcinoma	0	0	0	1	0	0.4444	
Adenoma/Carcinoma	0	0	0	2	2	0.0269	0.0941
Bone/Cranial and Facial Bones							
Osteoma	0	1	0	0	0	1.0000	1.0000
Brain							
Astrocytoma	0	0	1	1	0	0.3847	
Oligodendroglioma	0	1	0	0	0	1.0000	1.0000
Ear							
Squamous cell papilloma	1	0	0	0	0	1.0000	1.0000
Ear/Zymbal's Gland							
Squamous Cell Carcinoma	0	0	0	0	1	0.1437	0.2353
Esophagus							
Fibrosarcoma	0	1	0	0	0	1.0000	1.0000
Heart/Endocardium							
Schwannoma	0	1	0	1	0	0.5699	1.0000
Kidney							
Lipoma	0	0	0	1	0	0.3413	
Liver							
Adenoma	1	1	0	1	5	0.0015	0.0120
Adenoma/Carcinoma	2	4	6	2	12	0.0000	0.0000
Carcinoma	1	3	6	1	7	0.0044	0.0050
Hemangioma	0	1	0	0	0	1.0000	1.0000
Liver/Bile Duct							
Carcinoma	0	0	0	1	0	0.3413	
Lung/Mediastinum							
Fibrosarcoma	0	1	0	0	0	1.0000	1.0000
Lymph Node/Mesenteric							
Hemangioma	0	0	0	1	0	0.4444	
Mammary Gland							
Carcinoma	0	0	1	0	1	0.1423	0.2353
Pancreas/Acinus							
Adenoma	1	0	0	0	0	1.0000	1.0000
Pancreas/Islet							
Adenoma	0	4	3	0	0	0.9871	1.0000
Carcinoma	1	2	2	1	0	0.9125	1.0000
Parathyroid							
Adenoma	1	0	0	1	1	0.2284	0.4170
Peritoneum							
Lipoma	1	0	0	0	0	1.0000	1.0000
Pituitary							
Adenoma	27	20	21	25	19	0.3776	0.5205

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

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	High	p-values: Trend	Hi vs Cntrl
Primary Site Undetermined							
Leukemia	0	0	1	0	1	0.1502	0.2627
Lymphoma	1	0	0	0	1	0.2930	0.4571
Prostate							
Adenocarcinoma	0	0	1	0	0	0.5812	
Adenoma	1	0	0	0	0	1.0000	1.0000
Salivary Gland							
Adenoma	1	0	0	0	0	1.0000	1.0000
Skeletal Muscle							
Osteosarcoma	0	0	0	0	1	0.1757	0.2932
Skin							
Basal cell tumor	2	1	0	0	2	0.1586	0.3391
Carcinoma	0	0	0	0	1	0.1584	0.2667
Fibroma	1	0	1	0	0	0.7870	1.0000
Fibrosarcoma	0	0	1	0	0	0.5926	
Keratoacanthoma	1	1	1	1	0	0.7361	1.0000
Lipoma	0	0	0	0	1	0.3030	0.5882
Papilloma	1	0	1	1	0	0.6497	1.0000
Sarcoma	1	1	1	0	0	0.9172	1.0000
Sebaceous adenoma	0	1	0	0	0	1.0000	1.0000
Skin/Dermis							
Hemangiosarcoma	1	0	0	0	0	1.0000	1.0000
Small Intestine							
Leiomyoma	1	0	0	0	0	1.0000	1.0000
Leiomyosarcoma	0	0	0	0	1	0.1911	0.3219
Spleen							
Hemangiosarcoma	0	1	0	0	0	1.0000	1.0000
Stomach/Nonglandular mucosa							
Hemangiosarcoma	0	1	0	0	0	1.0000	1.0000
Papilloma	1	0	0	0	0	1.0000	1.0000
Systemic							
Hemangioma	0	1	0	1	0	0.7009	1.0000
Hemangiosarcoma	1	2	0	0	0	1.0000	1.0000
Hemangioma/-sarcoma	1	3	0	1	0	0.9326	1.0000
Testis							
Interstitial Cell Tumor	3	3	3	0	1	0.8747	0.9309
Thyroid/Follicular Cell							
Adenoma	0	0	1	0	1	0.1325	0.2353
Carcinoma	1	1	1	0	2	0.1601	0.3051
Thyroid/Parafollicular Cell							
Adenoma	3	6	5	5	2	0.7707	0.8197
Carcinoma	0	0	2	2	1	0.2390	0.2353
Urinary Bladder/Transitional E							
Carcinoma	0	0	0	0	1	0.1630	0.2655

Table A.4.4. Tumorgenicity in Female Rats

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	High	p-values: Trend	Hi vs Cntrl
Adrenal							
Pheochromocytoma	1	1	2	1	1	0.4589	0.6167
Adrenal/Cortex							
Adenoma	0	2	0	0	0	1.0000	1.0000
Carcinoma	0	0	2	0	0	0.6568	
Adenoma/Carcinoma	0	2	2	0	0	0.8986	1.0000
Brain							
Astrocytoma	1	1	1	1	0	0.7741	1.0000
Glioma	0	0	0	1	0	0.3777	
Granular cell tumor	0	0	0	1	1	0.1178	0.3441
Ear/Zymbal's Gland							
Squamous Cell Carcinoma	0	0	0	1	0	0.3910	
Heart/Endocardium							
Schwannoma	0	1	1	0	0	0.8490	1.0000
Liver							
Adenoma	1	0	2	3	3	0.0872	0.0920
Carcinoma	0	1	1	1	6	0.0008	0.0041
Adenoma/Carcinoma	1	1	3	4	9	0.0001	0.0004
Mammary Gland							
Adenoma	3	0	0	1	1	0.4814	0.7308
Carcinoma	17	16	10	10	5	0.9945	0.9982
Fibroadenoma	8	8	9	6	7	0.6075	0.6528
Ovary							
Sertoli Cell Tumor	1	1	0	1	0	0.8058	1.0000
Pancreas							
Schwannoma	0	0	1	0	0	0.6090	
Pancreas/Islet							
Adenoma	0	1	1	0	0	0.7930	1.0000
Parathyroid							
Adenoma	1	0	0	0	1	0.3692	0.5722
Peritoneum							
Paraganglioma	0	0	1	0	0	0.6090	
Pituitary							
Adenoma	34	42	35	28	19	0.9994	0.9997
Carcinoma	0	0	1	0	0	0.6139	
Primary Site Undetermined							
Histiocytic Sarcoma	0	1	0	1	1	0.2548	0.5530
Lymphoma	0	0	1	0	0	0.6092	
Skeletal Muscle							
Fibrosarcoma	0	0	1	0	0	0.6090	
Histiocytoma	0	0	1	0	0	0.6090	

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

Organ / Tumor	Con-trol1	Con-trol2	Low	Med-ium	High	p-values: Trend	Hi vs Cntrl
Skin							
Basal cell tumor	0	0	1	0	0	0.6090	
Fibroma	0	0	0	0	1	0.2051	0.3441
Fibrosarcoma	0	0	0	0	1	0.2051	0.3441
Histiocytoma	0	0	0	1	0	0.3910	
Keratoacanthoma	1	0	0	0	0	1.0000	1.0000
Osteosarcoma	1	0	0	0	0	1.0000	1.0000
Papilloma	1	1	0	0	0	1.0000	1.0000
Small Intestine							
Leiomyoma	0	0	1	0	0	0.6090	
Leiomyosarcoma	0	1	0	0	0	1.0000	1.0000
Stomach/Nonglandular mucosa							
Papilloma	0	0	0	0	1	0.1579	0.2308
Thyroid/Follicular Cell							
Adenoma	0	0	2	0	0	0.5170	
Carcinoma	0	0	0	1	0	0.3910	
Thyroid/Parafollicular Cell							
Adenoma	4	1	6	2	0	0.9664	1.0000
Carcinoma	2	0	1	2	1	0.4660	0.7227
Urinary Bladder							
Leiomyosarcoma	0	0	0	0	1	0.2051	0.3441
Uterus							
Cystadenoma	0	0	0	0	1	0.2051	0.3441
Fibroma	0	1	0	0	0	1.0000	1.0000
Granular cell tumor	1	0	0	0	0	1.0000	1.0000
Polyp	1	2	4	1	1	0.7109	0.7500
Sarcoma	0	0	0	0	1	0.2047	0.3398
Uterus/Cervix							
Polyp	1	2	0	1	0	0.8331	1.0000
Sarcoma	1	1	0	1	0	0.7785	1.0000
Vagina							
Granular cell tumor	0	2	0	0	0	1.0000	1.0000

Appendix 5. References

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

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

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

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

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

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

/s/

Steven Thomson
5/12/2006 03:12:29 PM
BIOMETRICS

Karl Lin
5/16/2006 11:01:59 AM
BIOMETRICS
Concur with review