CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

206038Orig1s000

STATISTICAL REVIEW(S)

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCES
OFFICE OF BIOSTATISTICS

Statistical Review and Evaluation

CLINICAL STUDIES

NDA:	206-038
Name of drug:	Orkambi (lumacaftor/ivacaftor)
Indication:	Treatment of cystic fibrosis in patients ≥ 12 years homozygous for F508del mutation
Applicant:	Vertex, Inc.
Dates:	Received: November 5, 2014 PDUFA: July 5, 2015
Review priority:	Priority – 8 month
Biometrics division:	Biometrics II
Statistical reviewer:	David Petullo, MS
Concurring reviewers:	Thomas Permutt, PhD
Medical division:	Pulmonary, Allergy and Rheumatology Products
Clinical team:	Robert Lim, MD Anthony Durmowicz, MD
Project manager:	Leila Hann
Keywords:	NDA review, clinical studies

1 INTRODUCTION

1.1 OVERVIEW

The applicant submitted the results from two phase 3 efficacy studies, VX12-809-103 (809-103) and VX12-809-104 (809-104) that evaluated placebo and two doses of lumacaftor/ivacaftor (LUM/IVA) in cystic fibrosis (CF) patients homozygous for the *F508del* mutation. These studies demonstrated a statistically significant treatment benefit in improving lung function (FEV_1). However, these studies did not evaluate either lumacaftor or ivacaftor as monotherapies. The Applicant's rationale for not including each component as monotherapy was that treatment with lumacaftor demonstrated a dose dependent decrease in percent predicted FEV_1 ($ppFEV_1$) in a phase 2 study, VX12-809-102 (809-102), and ivacaftor was shown to be ineffective in study VX08-770-104 (770-104). Study 770-104 was a phase 2 study that was submitted and reviewed in the application for ivacaftor, NDA 203-188. Based on the results from this study, the current label indicates in the *Limitations of Use* section that ivacaftor is "not effective in patients with CF who are homozygous for the F508del mutation in the CFTR gene."

The Division agreed that evaluation of lumacaftor as a monotherapy would not be required as it is unlikely to be developed for use as a single ingredient therapy and demonstrated a dose-dependent decrease in $ppFEV_1$. However, ivacaftor did not demonstrate such a decrease in $ppFEV_1$. To evaluate the contribution of lumacaftor, I discuss the results from study 770-104 with respect to the results from studies 809-103 and 809-104 using a non-inferiority (NI) approach. In studies 809-103 and 809-104, this approach considers the dose indicated on the proposed label, LUM 400mg/IVA 250mg q12h, and changes in $ppFEV_1$ and exacerbation rates. Although all studies evaluated changes in BMI and CFQ-R, these endpoints were not included as they failed to show substantial evidence of a treatment effect in in studies 809-103 and 809-104. Changes in $ppFEV_1$ and exacerbation rates were significantly better than placebo in these studies. Adjustments for multiplicity were not considered: a p-value < 0.05 was considered significant.

Studies 809-103 and 809-104 did not evaluate changes in sweat chloride, a diagnostic marker of CF and indicator of CFTR ion channel function. The results from study 809-102 were evaluated to explore changes in sweat chloride following 4 weeks of treatment with lumacaftor monotherapy followed by 4 weeks of treatment with LUM/IVA.

The Pulmonary, Allergy, Drugs Advisory Committee convened on May 12, 2015 to discuss this application and whether they believed sufficient evidence was present to evaluate the contribution of lumacaftor to the combination product and overall approval of Orkambi. Even though the majority of Committee members agreed that there was not sufficient evidence to establish that the combination product was any better than ivacaftor monotherapy, they voted 12–1 to recommend approval of the combination product.

1.2 DATA SOURCES

All data was supplied electronically by the applicant as SAS transport files and was of sufficient quality to allow a thorough review. The data can be found at the following location in the CDER electronic document room (EDR):

\\cdsesub1\evsprod\NDA206038\0002\m5\datasets

\\cdsesub1\evsprod\NDA203188\0002\m5\datasets

2 CONTRIBUTION OF LUMACAFTOR

The applicant's rationale for not evaluating ivacaftor or lumacaftor as monotherapy is given below.

> A lumacaftor monotherapy arm was not included in Studies 103 and 104 because results from Study 102 showed a dose-dependent decline in percent predicted FEV1 during treatment with lumacaftor monotherapy. This decline was statistically significant at the highest lumacaftor dose tested (400 mg q12h, within-group analysis). Given the lack of efficacy of lumacaftor monotherapy in clinical studies, coupled with a low response in vitro to lumacaftor alone in airway epithelial cells from patients homozygous for the F508del-CFTR mutation, further clinical evaluation of lumacaftor monotherapy was considered unlikely to reveal significant benefit.
>
> An ivacaftor monotherapy arm was not included in Studies 103 and 104 because evaluating the overall results from Study 770-104 had previously demonstrated that there was no clinically meaningful benefit after 16 weeks of treatment with ivacaftor monotherapy (150 mg q12h) in subjects homozygous for the F508del-CFTR mutation.

Even though the Agency agreed with the rationale for not evaluating the contribution of ivacaftor or lumacaftor to the combination product, there is now a concern regarding the contribution of lumacaftor: i.e., Does the combination product provides a treatment benefit over ivacaftor monotherapy? The Applicant's argument in our view is essentially identical to a non-inferiority (NI) argument but is missing an essential piece. The Applicant concluded that LUM/IVA is superior to ivacaftor by making the following argument:

1. Placebo was shown to be similar or non-inferior to ivacaftor (Study 770-104);
2. LUM/IVA was superior to placebo (Studies 809-103 and 809-104);
3. Therefore, LUM/IVA is better than ivacaftor since LUM/IVA beat placebo and placebo was non-inferior to ivacaftor.

The absence of a statistically significant difference between ivacaftor alone and placebo does not in itself establish that ivacaftor is similar enough to placebo to sustain this argument. Study 770-104 does, however, provide an upper confidence bound on the difference between ivacaftor and placebo, so that it permits the inferiority of placebo to

ivacaftor to be assessed. As I will show, this bound is not tight enough for the purpose. That is, the combined results of Study 770-104 with 809-103 and 809-104 do not show that the effect LUM/IVA is more than that of ivacaftor alone.

To evaluate the contribution of lumacaftor to LUM/IVA, I continued with the NI approach to test if LUM/IVA was superior to ivacaftor with respect to ppFEV1 and pulmonary exacerbations utilizing the synthesis method. This approach required an assessment of the constancy assumption: i.e., were the studies similar in design, patient population, standard of care, and so forth. The synthesis method, which does not require a NI margin, allowed a comparison of LUM/IVA to ivacaftor by combining the variance across studies. A 95% confidence (CI) for the difference between ivacaftor and LUM/IVA was derived. If this CI excluded 0 (1 for exacerbations), one could conclude with 95% confidence that LUM/IVA was superior to placebo if one is willing to accept that the effect of ivacaftor in study 770-104 was similar to that in studies 809-103 and 809-104.

Below I discuss the study design, endpoints, statistical analysis, and results from study 770-104. For a detailed review of studies 809-103 and 809-104, the reader is referred to the clinical and statistical review of these studies.

2.1 STUDY DESIGN

Study 770-104 was a randomized, double-blind, placebo-controlled trial that evaluated the safety and efficacy of ivacaftor in subjects 12 years of age and older with CF homozygous for the *F508del CFTR* mutation. This study was conducted in two parts, A and B. Part A was a 16-week, double-blind, placebo-controlled study where eligible subjects were randomized in 4:1 ratio to ivacaftor 150 mg or placebo administered every 12 hours (q12h). No formal sample size and power calculations were performed for this study with respect to efficacy. A sample size of 120 subjects was deemed adequate to evaluate the safety of the ivacaftor in patients homozygous for the *F508del* mutation. Subjects who completed 16 weeks of study drug treatment in Part A, and met a pre-defined responder criterion could elect to participate in Part B. The protocol-defined responder criteria were established based on changes in FEV_1 and sweat chloride, given the desired clinical benefit of ivacaftor (improvement in lung function) and its mechanism of action. This review focuses on Part A of study 770-104.

FEV_1 was measured at baseline, day 15, Week 8, and Week 16. Pulmonary exacerbations were defined as new, or changed, antibiotic therapy (IV, inhaled, or oral) for any four or more of the following signs or symptoms: change in sputum, new or increased hemoptysis, increased cough, increased dyspnea, malaise, fatigue, or lethargy, temperature above 38°C, anorexia or weight loss, sinus pain or tenderness, change in sinus discharge, change in physical examination of the chest, decrease in pulmonary function by 10%, radiographic changes indicative of pulmonary infection. This definition of an exacerbation was the same in studies 809-103 and 809-104.

2.2 STATISTICAL METHODOLOGIES

Changes from baseline at Week 16 in $ppFEV_1$ (absolute and relative), were compared between ivacaftor and placebo using an ANCOVA model with treatment and baseline value. A crude rate of pulmonary exacerbation (number of events/days on study) was determined for each treatment arm and the rate ratio between placebo and ivacaftor was obtained.

To evaluate the contribution of lumacaftor, I explored a NI approach after examining key issues such as study duration, inclusion and exclusion criteria, standard of care, and so forth. Even though there were differences noted between the ivacaftor study and the LUM/IVA studies, I determined that the studies were similar enough in design to continue with the NI approach even though the constancy assumption may not be valid. The applicant's argument also relied on this untestable assumption.

To obtain an estimate of the treatment effect with respect to $ppFEV_1$ and rate of pulmonary exacerbations, the results from studies 809-103 and 809-104 were integrated. Next, 95% CIs for the difference from placebo were calculated for ivacaftor (study 770-104) and LUM/IVA. The synthesis method was used to combine the variance from study 770-104 with the integrated results from studies 809-103 and 809-104. If the 95% CI for the difference between LUM/IVA and ivacaftor does not contain zero, superiority was established.

2.3 PATIENT DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Demographic characteristics and patient disposition for randomized subjects randomized are shown in Table 1.

Table 1. Demographics for Study 770-104

Characteristic	placebo	ivacaftor 150 mg
Number of Patients	28	112
Age in years		
Mean (SD)	25.0 (8.4)	22.8 (10.3)
Median	24	19.5
[range]	[12, 39]	[12, 52]
Gender, n (%)		
Female	16 (57)	58 (52)
Male	12 (43)	54 (48)
Race, n (%)		
Caucasian	28 (100)	111 (99)
Black	-	1 (1)

Source: Reviewer

The majority of subjects completed 16-weeks of treatment (Table 2). TA total of 2 placebo subjects and 8 ivacaftor subjects did not complete Part A.

Table 2. Subject disposition for Study 770-104

Disposition	placebo	ivacaftor
Randomized	28	112
Completed Part A	26	104
discontinued	2	8
Reason for Discontinuation		
adverse event	2	3
lost to follow-up	-	1
noncompliance	-	2
prohibited medication	-	1
study termination	-	1

Source: Reviewer

2.4 RESULTS

Study 770-104 was previously reviewed under NDA 203-188 and failed to show a significant treatment benefit for ivacaftor with respect to $ppFEV_1$, CFQ-R, BMI, and rate of pulmonary exacerbations. Changes in sweat chloride were significantly different from placebo without any adjustments for multiplicity. The 95% CI for difference from placebo for the change from baseline through Week 16 is shown for each endpoint in Table 3. Exacerbation is presented as a rate ratio of ivacaftor and placebo.

Table 3. Efficacy results from study 770-104

Endpoint	difference from placebo	95% CI
ΔppFEV1	1.7[a]	[-0.6, 4.1]
Δ sweat chloride	-2.9[a]	[-5.6, -0.2]
Δ CFQ-R	1.3[a]	[-2.9, 5.6]
Δ BMI	-0.07[b]	[-0.4, 0.2]
Exacerbation (rate ratio)	0.68[c]	[0.33, 1.4]

Source: Reviewer
a: MMRM with baseline, age, visit, treatment, and visit*treatment
b: LME with treatment, visit, age, baseline $ppFEV_1$
#: Poisson regression with age, baseline $ppFEV_1$, visit, treatment, and visit*treatment

Considering the constancy assumption, the similarity of the trials was examined. The main differences between study 770-104 and studies 809-103 and 809-104 were duration of treatment and inclusion criteria. Study 770-104 consisted of 16 weeks of double-blind treatment whereas studies 809-103 and 809-104 were 24 weeks. To adjust for the differences in treatment duration, for changes in $ppFEV_1$, the change from baseline at Week 16 was examined rather than through 16 weeks, i.e., a repeated measures analysis. For exacerbation, annualized crude rates were considered without an adjustment for treatment duration. With respect to baseline lung function, subjects with a baseline $ppFEV_1$ greater than 90% were excluded. There were 8 placebo subjects and 38 ivacaftor subjects whose baseline $ppFEV_1$ was greater than 90%. Additionally one subject in study 809-103 and five subjects in study 809-104had a baseline $ppFEV_1$ greater than 90%. The results using all patients regardless of baseline lung function is also reported as it would represent all randomized and treated subjects.

Even though there were differences in the demographics and baseline characteristics across the three studies I continued with the NI approach keeping in mind that the constancy assumption may not be valid. The reader is referred to the clinical and statistical review of studies 770-104, 809-103, and 809-104 for details of the demographics and baseline characteristics.

$ppFEV_1$: The results for lung function are presented in terms of $ppFEV_1$ at baseline and change from baseline at Week 16 (absolute and relative) for each study in Table 4. From studies 809-103 and 809-104, only the results from the dose that is indicated on the proposed label, LUM 400mg/IVA 250mg q12h were presented.

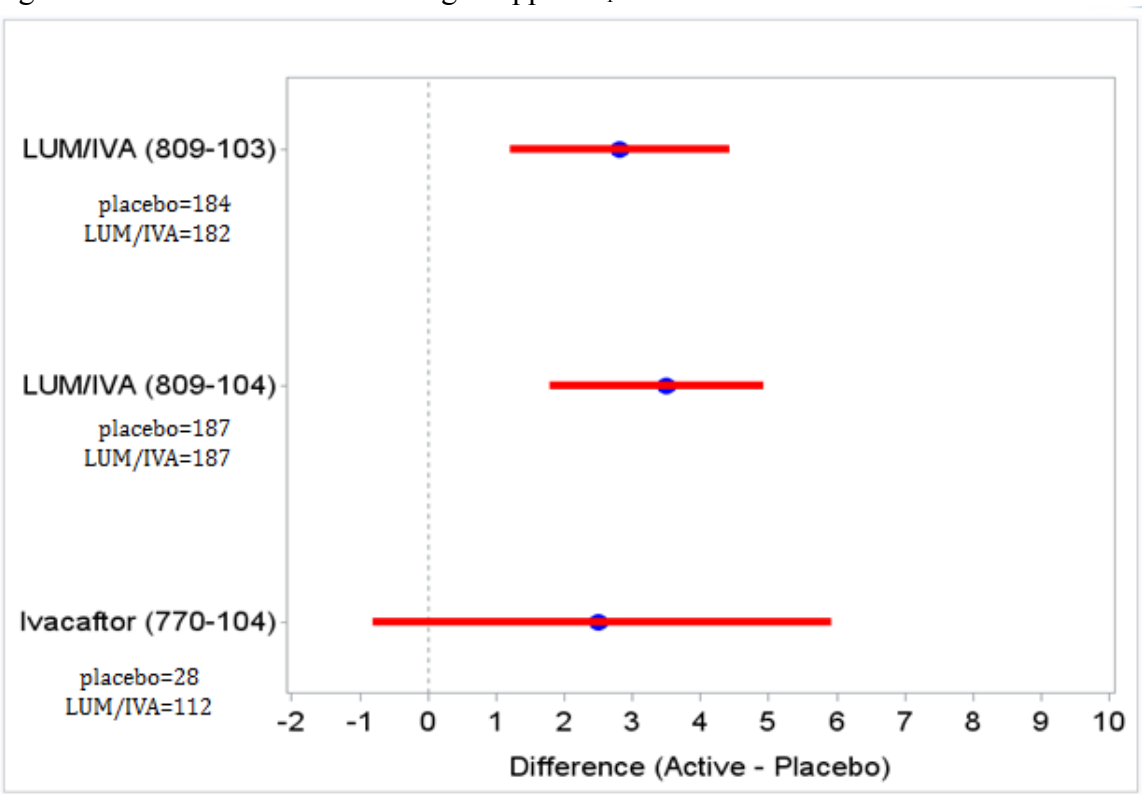
Table 4. Summary of $ppFEV_1$ at week 16 for all randomized and treated subjects

Study	Time Point	$ppFEV_1$, LSMEAN* (SE)		
		placebo	Ivacaftor (150 mg q12h)	LUM 400mg / IVA 250mg q12h
770-104	Baseline	73.2 (4.5)	76.9 (2.2)	-
	Absolute Change*	-0.3 (1.5)	2.2 (0.8)	-
	Relative change#*	-0.4 (2.1)	3.2 (1.1)	-
809-103	Baseline	60.3 (1.0)	-	60.5 (1.1)
	Absolute Change	-0.2 (0.6)	-	2.6 (0.6)
	Relative change#	0.3 (1.0)	-	4.7 (1.0)
809-104	Baseline	60.2 (1.0)	-	60.3 (1.1)
	Absolute Change	-0.7 (0.6)	-	2.8 (0.6)
	Relative change#	-0.7 (1.0)	-	5.4 (1.0)

Source: Reviewer
*ANCOVA with baseline $ppFEV_1$, # Relative change is define as % change from baseline

The difference from placebo for change in $ppFEV_1$ for ivacaftor and LUM/IVA for all randomized subjects regardless of baseline lung function is presented in Figure 1 along with the associated 95% CI.

Figure 1. Treatment effect for change in $ppFEV_1$



Source: Reviewer

The data from study 770-104 was combined or synthesized with the results from the integration of studies 809-103 and 809-104. The 95% CI for the difference between LUM/IVA and ivacaftor is shown in Table 5. Results are presented with respect to baseline $ppFEV_1$, ≥ 40% or between 40 and 90%.

Table 5. Comparison of $ppFEV_1$ for LUM/IVA and ivacaftor by baseline $ppFEV_1$

Baseline $ppFEV_1$	Study	LSMEAN* [95% CI]	SE
40-90%	LUM/IVA (integrated)	3.2 [2.1, 4.3]	0.6
	770-104	2.6 [-1.1, 6.4]	1.9
	Combo-Mono (synthesized)	0.6 [-3.3, 4.5]	2.0
≥40%	LUM/IVA (integrated)	3.2 [2.1, 4.3]	0.6
	770-104	2.5 [-0.8, 5.9]	1.7
	Combo-Mono (synthesized)	0.7 [-2.8, 4.1]	1.8

Source: Reviewer
*ANCOVA with treatment and baseline $ppFEV_1$

Superiority was not established. Inclusion or exclusion of subjects based on baseline lung function was irrelevant.

Pulmonary Exacerbations: Crude exacerbation rates for the integrated LUM/IVA studies and study 770-104 are presented in Table 4. The definition of an exacerbation was identical in all studies and was defined as new or changed antibiotic therapy (IV, inhaled, or oral) for any four or more of the following signs or symptoms. As results were similar with respect to baseline lung function, results are presented for all randomized and treated subjects and are shown in Table 6.

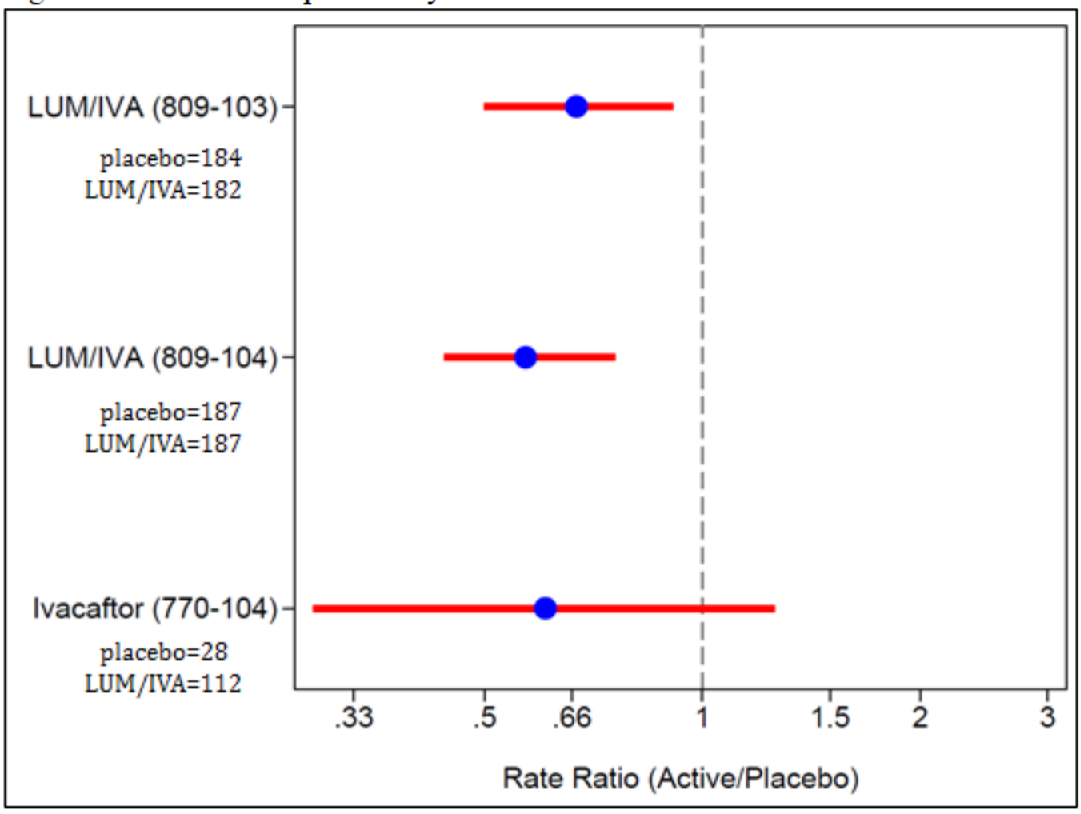
Table 6. Summary of pulmonary exacerbations

Study	Statistics	Exacerbations		
		placebo	Ivacaftor (150 mg q12h)	LUM 400mg / IVA 250mg q12h
770-104	n	28	112	-
	days on study	3038	12504	-
	Annual rate	1.2	0.73	-
	Rate ratio	-	0.61	-
	SE		0.37	-
Integrated*	n	371	-	369
	days on study	62427	-	61,057
	Annual rate	1.5	-	0.91
	Rate ratio	-	-	0.62
	SE	-	-	0.1

Source: Reviewer
*Studies 809-103 and 809-104

The rate ratios and associated 95% CIs are shown Figure 2. A ratio of one would indicate no difference in exacerbation rates.

Figure 2. Rate ratio for pulmonary exacerbations



Source: Reviewer

A direct comparison of LUM/IVA to ivacaftor is presented via the synthesis method in Table 7. Results are summarized based on baseline $ppFEV_1$, 40% or greater and between 40% and 90%. When I considered baseline $ppFEV_1$ between 40% and 90%, 10 subjects with 10 events were excluded from study 770-104. No events were excluded from the LUM/IVA studies. However, days on study for those subjects that had baseline $ppFEV_1$ greater than 90% were not counted when determining exacerbation rates.

Table 7. Rate ratio of pulmonary exacerbations

Baseline $ppFEV_1$	Study	Exacerbations Rate Ratio [95% CI]	SE
40-90%	LUM/IVA (integrated)	0.62 [0.51, 0.76]	0.10
	770-104	0.56 [0.24, 1.31]	0.43
	Combo/Mono (synthesized)	1.11 [0.47, 2.62]	0.44
≥40%	LUM/IVA (integrated)	0.62 [0.51, 0.76]	0.06
	770-104	0.61 [0.33, 1.37]	0.37
	Combo/Mono (synthesized)	1.02 [0.48, 2.18]	0.39

Source: Reviewer

Again, as observed with change in $ppFEV_1$, superiority was not established and baseline lung function was irrelevant.

2.5 DISCUSSION

The Applicant's claim that (b) (4) is misleading. (b) (4) To correct this, I used an approach that assumed the effect of ivacaftor in study 770-104 would be similar in studies 809-103 and 809-104 not that ivacaftor was similar to placebo. Even though I do not necessarily believe the constancy assumption was met, but even if I did, the results from these analyses could not conclude with any level of confidence that LUM/IVA was significantly different from ivacaftor with respect to changes in $ppFEV_1$ and pulmonary exacerbations. Exclusion of subjects with greater than 90% $ppFEV_1$ at baseline did not affect the conclusions.

On February 23 an information request was sent to the sponsor requesting an evaluation of the contribution of lumacaftor to LUM/IVA. Below is an excerpt from the response that was received.

1. The nature of the molecular defect caused by the *F508del-CFTR* mutation is well established and LUM is essential to address the underlying cause of disease in patients homozygous for the *F508del-CFTR* mutation.

2. Nonclinical data quantitate the contribution of each drug to the improvement in F508del-CFTR function, and show that there is minimal effect of IVA alone while superior improvement in F508del-CFTR function is provided by the combination of LUM and IVA compared to either agent alone.

3. The improvements in F508del-CFTR function in vitro translate to the sweat chloride response in subjects homozygous for the *F508del-CFTR* mutation, and confirm that superior improvement is provided by the combination of LUM and IVA compared to either agent alone.

4. Clinical evidence demonstrates that LUM/IVA combination therapy is highly efficacious and clinically superior to IVA monotherapy in homozygous *F508del-CFTR* subjects, confirming that LUM is an essential component of the combination product.

A robust Phase 3 clinical program demonstrated rapid, consistent, and sustained improvements in respiratory and systemic parameters with LUM/IVA combination therapy, notably including marked reductions in severe pulmonary exacerbations. LUM/IVA was well-tolerated, with a favorable safety profile in more than 1000 subjects. This positive clinical benefit/risk profile supports approval of the LUM/IVA combination therapy in patients age 12 years and older who are homozygous for the *F508del* mutation in the *CFTR* gene. In contrast, as indicated in the Kalydeco label, IVA monotherapy evaluated in this population did not show a consistent and meaningful clinical benefit.

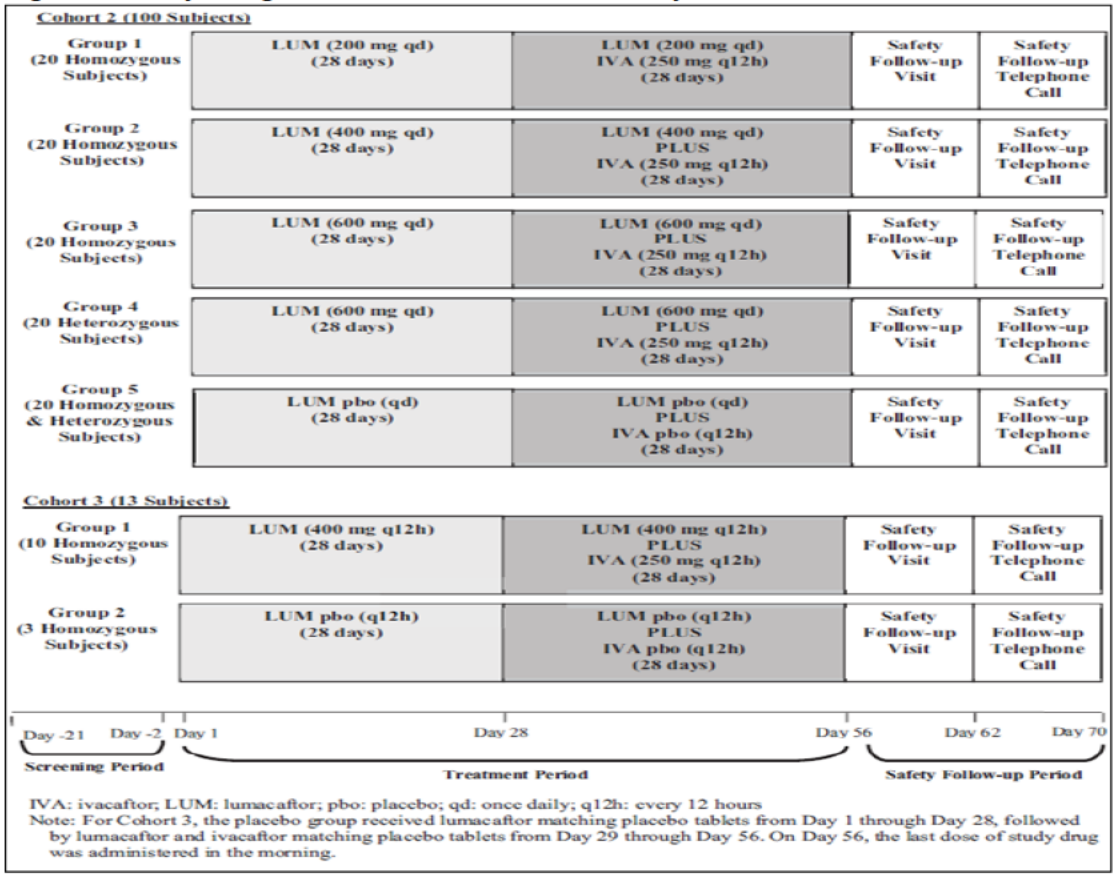
The applicant also provided the results from an integrated analysis of studies 770-104, 809-103, and 809-104 noting the limitations of such an analysis (results not shown). This analysis did not demonstrate a significant difference between LUM/IVA and ivacaftor.

3 SWEAT CHLORIDE

3.1 STUDY DESIGN

Study 809-102 was a randomized, double-blind, placebo-controlled, dose-ranging study that evaluated lumacaftor monotherapy and LUM/IVA in subjects that had the *F508del* mutation in the *CFTR* gene. This study was conducted in four different cohorts; however, this review focuses on the homozygous subjects from cohorts 2 and 3, where subjects received lumacaftor for 28 days followed by LUM/IVA for an additional 28 days. A schematic of the study design for cohorts 2 and 3 is shown in Figure 3. Sweat chloride was measured before study drug administration at baseline and on days 28 and 56. An additional measurement was taken 4 hours after study drug administration on days 28 and 56.

Figure 3. Study design for Cohorts 1and 2 from study 809-102



Source: modified from figures 9-2 and 9-3 of applicant's CSR

3.2 STATISTICAL METHODOLOGIES

Sweat chloride (mmol/L) is presented as the mean and standard deviation at baseline, Day 28, change from baseline at Day 28, Day 56, and the change from baseline at Day 56. The change from baseline at Days 28 and 56 used sweat chloride values measured at treatment administration and 4 hours after treatment administration. The change in sweat chloride was compared to placebo for each dosing regimen of active drug using an ANCOVA model with treatment and baseline sweat chloride. In this analysis, the placebo groups for cohorts 2 and 3 were pooled as the changes in sweat chloride appeared similar. Missing data was minimal and was not imputed in these analyses.

3.3 PATIENT DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Demographic characteristics and patient disposition for homozygous subjects randomized in cohorts 2 and 3 are shown in Table 8.

Table 8. Demographics for all randomized subjects homozygous for F508 del mutation

Characteristic	Placebo (pooled)	lumacaftor/ivacaftor (mg)[c]			
		200[a]250[b]	400[a]/250[b]	600[a]/250[b]	400[b]/250[b]
Number of Patients	21	23	21	21	11
Age in years					
Mean (SD)	30.4 (11.5)	28.1 (9.0)	27.1 (6.8)	26.7 (6.5)	25.5 (6.7)
Median	28	26	26	26	25
[range]	[18,63]	[18, 52]	[18, 42]	[18, 42]	[18, 37]
Gender, n (%)					
Female	8 (38)	11 (48)	9 (43)	11 (52)	5 (45)
Male	13 (62)	12 (52)	12 (57)	10 (48)	6 (54)
Race, n (%)					
Caucasian	21 (100)	23 (100)	21 (100)	21 (100)	11 (100)
Black	-	-	-	-	-
Other	-	-	-	-	-

Source: Reviewer
a: once daily dosing, b: twice daily dosing, c: during weeks 1-4, subjects received lumacaftor, in weeks 5-8, subjects received lumacaftor/ivacaftor.

In cohort 2, 6 subjects discontinued dosing due to an adverse event: 2 in the 200/250 treatment arm, 1 in the 400/250 arm, and 3 in the 600/250 arm. In cohort 3, 2 subjects discontinued due to an adverse event, one in each treatment arm. However, only one subject in cohort 2 did not complete the safety follow-up period.

3.4 RESULTS

A summary of sweat chloride results is presented in Table 9. Results are summarized according to when sweat chloride was measured.

Table 9. Summary of sweat chloride data for homozygous subjects in Cohorts 2 and 3

SWCl measured	Time	Sweat Chloride (mmol/L), mean (stdev) Pooled Placebo	lumacaftor/ivacaftor (mg)[c] 200[a]/250[b]	400[a]/250[b]	600[a]/250[b]	400[b]/250[b]
At dosing	Baseline	97.5 (8.8)	97.1 (9.8)	98.2 (7.1)	98.8 (11.9)	102.4 (8.9)
	Δ Day 28	0.6 (7.7)	-4.4 (6.8)	-8.1 (7.6)	-6.0 (11.0)	-9.3 (9.2)
	Δ Day 56	0.2 (9.3)	-3.9 (9.6)	-8.9 (11.4)	-8.9 (10.2)	-12.2 (6.6)
4-hours post-dose	Baseline	97.5 (8.8)	97.1 (9.8)	98.2 (7.1)	98.8 (11.9)	102.4 (8.9)
	Δ Day 28	3.7 (7.7)	-3.3 (9.4)	-7.1 (14.3)	-8.4 (10.8)	-2.6 (14.3)
	Δ Day 56	3.2 (10.9)	0.2 (9.0)	-6.1 (12.5)	-6.4 (11.5)	-3.4 (7.8)

Source: Reviewer
a: once daily dosing, b: twice daily dosing, c: during weeks 1-4, subjects received lumacaftor, during weeks 5-8, subjects received lumacaftor/ivacaftor

Regardless of when sweat chloride was measured, either at time of study drug administration or 4 hours after administration, there was a decrease in sweat chloride irrespective of dose. Difference from placebo is shown in Table 10.

Table 10. Difference from placebo for change in sweat chloride

measured	Day	Difference from placebo for change in SWCL (mmol/L) , LSMEAN[d] [95% CI] lumacaftor/ivacaftor (mg)[c] 200[a]/250[b]	400[a]/250[b]	600[a]/250[b]	400[b]/250[b]
At dosing	28	-5.5 [-10.5,-0.5]	-8.8 [-13.9, -3.7]	-6.7 [-11.8. -1.7]	-8.9 [-15.2, -2.6]
	56	-4.7 [-10.7, 1.3]	-9.5 [-15.5, -3.5]	-9.2 [-15.3, -3.1]	-10.7 [-18.5, -2.9]
4-hours Post-dose	28	-7.0 [-13.8, -0.2]	-6.3 [-15.7, 3.1]	-12.1 [-19.1, -5.1]	-4.2 [-13.5, 5.0]
	56	-3.5 [-10.1, 3.0]	-9.5 [-16.1, -2.9]	-9.6 [-16.2, -2.9]	-5.0 [-13.2,3.2]

Source: Reviewer
a: once daily dosing, b: twice daily dosing, c: during weeks 1-4, subjects received lumacaftor, in weeks 5-8, subjects received lumacaftor/ivacaftor, d: ANCOVA with treatment and baseline lung function

3.5 DISCUSSION

There was a decrease in sweat chloride following four weeks of treatment with lumacaftor regardless of dose. However, the effect observed after an additional four weeks of treatment with LUM/IVA depends on when sweat chloride was measured. If measured at the time of treatment administration, there was some added benefit; i.e., there was a numerical decrease in sweat chloride following an additional four weeks of treatment with LUM/IVA. However, if sweat chloride was measured four hours after

study drug administration, there was no additional benefit, in fact in most cases the sweat chloride increased although it did not return to baseline levels. However, these decreases observed for sweat chloride, approximately10 mmol/L, were small especially in the context of the sweat chloride response observed for ivacaftor in the *G551D* and *R117H* mutations, approximately 50 and 24 mmol/L, respectively.

4 CONCLUSION

The contribution of lumacaftor to the efficacy of the proposed combination product has not been shown. In addition, the Applicant reported that the results from study 770-104 demonstrated that ivacaftor provides no clinically meaningful benefit. However, the estimated effect of LUM/IVA on $ppFEV_1$ was 2–3% which was similar to the effect noted for ivacaftor (estimate: 1.7%; 95% CI: -0.6%, 4.1%).

Treatment with lumacaftor for four weeks produced a small decrease in sweat chloride that was maintained after an additional four weeks of treatment with LUM/IVA. There was some variability in the results depending on when sweat chloride was measured, either at dosing or 4 hours after doing. Regardless, the mean decreased observed at Week 16 was still numerically lower than mean response at baseline. However, the decreases observed for sweat chloride were small when compared to the sweat chloride response noted for ivacaftor in the *G551D* and *R117H* mutations.

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

/s/

DAVID M PETULLO
06/02/2015

THOMAS J PERMUTT
06/02/2015
I concur.

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

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number:	NDA 206-038
Drug Name:	Orkambi
Indication(s):	Cystic Fibrosis (CF)
Applicant:	Vertex
Date(s):	Received: November 5, 2014 PDUFA: July 5, 2015
Review Priority:	Priority – 8 month
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Lan Zeng, M.S.
Concurring Reviewers:	David Petullo, M.S.
Medical Division:	Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Robert Lim, M.D., Medical Reviewer Anthony Durmowicz, M.D., Medical Team Leader
Project Manager:	Leila Hann

Keywords: fixed dose combination, hierarchical testing, mixed models, sample size

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Vertex Pharmaceuticals proposes Orkambi, a fixed dose combination (FDC) of lumacaftor 400mg and ivacaftor 125mg oral tablet twice daily (LUM 400mg/IVA 250mg q12h) for the treatment of cystic fibrosis (CF) in patients12 years and older who are homozygous for the *F508del* mutation in the *CFTR* gene. Efficacy and safety of this FDC product and a higher dose combination (LUM 600mg qd /IVA 250mg q12h) were examined in two phase 3 trials.

The submission demonstrated benefits of both dosing regimens over placebo in terms of pulmonary lung function. Two replicated randomized parallel arm trials, VX12-809-103 (809-103) and VX12-809-104 (809-104), showed that both doses of Orkambi provided statistically significant benefits over placebo with regard to the primary endpoint, absolute change from baseline in percent predicted forced expiratory volume in 1 second ($ppFEV_1$) at Week 24, assessed as the average of the treatment effects at Week 16 and at Week 24. In study 809-103, the average treatment effect over placebo was 4.0% for LUM 600mg qd/IVA 250mg q12h and 2.6% for LUM 400mg /IVA 250mg q12h, respectively. In study 809-104, the average treatment effect above placebo was 2.6% for LUM 600mg qd/IVA 250mg q12h and 3.0% for LUM 400mg /IVA 250mg q12h, respectively. The results were consistent regardless of demographic subgroups, disease severity, or how change in $ppFEV_1$ was determined. Missing data was minimal and was not a concern.

In both studies, treatment with Orkambi resulted in improvements favoring active treatment over placebo for various key secondary endpoints: relative change from baseline in $ppFEV_1$ at Week 24 (assessed as the average of the treatment effects at Week 16 and at Week 24), absolute change from baseline in body mass index (BMI) at Week 24, absolute change from baseline in the Cystic Fibrosis Questionnaire-Revised (CFQ-R) respiratory domain at Week 24, response defined as ≥5% increase in relative change from baseline in $ppFEV_1$, and number of pulmonary exacerbations through Week 24. However, based on a pre-specified hierarchical testing strategy, only the relative change from baseline in $ppFEV_1$ provided replicated evidence demonstrating efficacy for both dosing regimens. While differences were noted in response rate in $ppFEV_1$and pulmonary exacerbation rates, these were not considered statistically significant as the endpoints tested before them failed to reach significance.

Regardless of statistical evidence, the clinical benefit of lumacaftor and ivacaftor combination treatment remains to be understood. Specifically for the claimed dose of LUM 400 mg q12h/ IVA 250 mg q12h, the average benefit was 2.6% to 3.0% over placebo. Furthermore, the Phase 3 studies by design did not evaluate contribution of each constitute component to the combination therapy. The reader is referred to the statistical review by Mr. David Petullo for additional discussion on these issues.

The Pulmonary and Allergy Drugs Advisory Committee convened on May 12, 2015 and discussed whether sufficient evidence was present to evaluate the contribution of lumacaftor to the combination product. Even though the majority of Committee members agreed that there was not enough evidence to establish that the combination product was any better than ivacaftor monotherapy, they voted 12–1 to recommend approval of the combination product.

2 INTRODUCTION

2.1 OVERVIEW

2.1.1 Drug Class and Indication

The current application evaluates Orkambi, a FDC of lumacaftor (LUM) and ivacaftor (IVA) for the treatment of CF) in patients age 12 years and older who are homozygous for the *F508del* mutation in the *CFTR* gene. Two dose levels of lumacaftor and ivacaftor combination (LUM/IVA) are studied, lumacaftor 400mg and ivacaftor 125mg twice daily (LUM 400mg/IVA 250mg q12h) as well as lumacaftor 600mg and ivacaftor 125mg twice daily (LUM 600mg qd /IVA 250mg q12h).

2.1.2 History of Drug Development

The clinical development program for LUM and IVA was introduced to the Division of Pulmonary, Allergy, and Rheumatology Products in 2007 under IND 79,521. The program consists of 17 clinical trials including one phase 2 trial (Study VX08-770-104, referred to as 770-104) which evaluated the effect of IVA monotherapy in subjects homozygous for the *F508del-CFTR* mutation and two phase 2 trials (Study VX08-809-101 and Study VX09-809-102 Cohorts 1, 2, and 3) that evaluated the effect of LUM monotherapy in this population. Study VX09-809-102 also assessed the efficacy of LUM/IVA combination in CF subjects with homozygous *F508del-CFTR* mutation. To demonstrate efficacy of LUM/IVA for the treatment of CF patients homozygous for the *F508del CFTR* gene, the applicant submitted the results from two replicate, 24-week, randomized, double-blind, and parallel group phase 3 trials, studies 809-103 and 809-104. Subjects in studies 809-103 and 809-104 who completed treatment were eligible to enroll in the long-term safety and efficacy rollover study (Study VX12-809-105) to receive active treatment for up to an additional 96 weeks.

The applicant had several interactions with the Agency, including two End-of-Phase 2 meetings held on November 2, 2012 and February 12, 2013, a Type B meeting held on January 8, 2014, and a Pre-NDA meeting held on August 12, 2014. Pertinent parts of the statistical portion of these meetings are summarized herein.

The Division and the applicant agreed that for the pivotal Phase 3 studies,

- Lumacaftor monotherapy and ivacaftor as monotherapy was not required for the phase 3 studies;
- Data through the end of treatment (24 weeks) should be submitted for review;
- The primary efficacy endpoint was absolute change from baseline in percent predicted FEV_1 ($ppFEV_1$). Relative change in $ppFEV_1$ could be a key secondary endpoint;
- The proposed mixed effects model with repeated measures (MMRM) was reasonable as long as the amount of missing data with respect to the primary endpoint was minimal;
- The proposed hierarchical testing strategy with Bonferroni correction was adequate for the control of Type I error.

The Division strongly recommended including assessments of sweat chloride before and after treatment in studies 809-103 and 809-104. The applicant indicated that it was not feasible to obtain baseline sweat chloride values for patients since the phase 3 studies were fully enrolled.

Furthermore, the Division noted that studies 809-103 and 809-104 had high statistical power to detect small effects with respect to the primary efficacy endpoint, $ppFEV_1$. In an earlier phase 2 study (770-104) a mean absolute change from baseline in $ppFEV_1$ relative to placebo was 1.7%, which was considered as not having a clinically meaningful treatment effect (see current label for ivacaftor). Review of the efficacy of LUM/IVA would consider not only the statistical significance of a treatment effect, but also the clinical importance for both primary and secondary endpoints.

2.1.3 Current Submission

The current submission contains the results from two phase 3 trials, 809-103 and 809-104. In addition the applicant submitted the results from a phase 2 study (770-104) that evaluated ivacaftor as monotherapy. Study 770-104 had previously been submitted under NDA 203-188 and was reviewed by Dr. David Hoberman in 2012. The applicant was requested to submit study 770-104 as it is the only study that evaluated ivacaftor monotherapy.

This review will focus on the phase 3 studies that evaluated the efficacy of the combination product LUM/IVA. The phase 2 study will be evaluated in a separate review by David Petullo that considers the contribution of lumacaftor to the combination product.

2.2 DATA SOURCES

The applicant submitted NDA 206-038 including clinical study reports, protocols, statistical analysis plan, and all referenced literature to the Agency. The data and final study reports for the electronic submission were archived under the network path location

\\Cdsesub1\evsprod\NDA206038\0002.

3 STATISICAL EVALUATION

3.1 DATA AND ANALYSIS QUALITY

In general, the electronic data submitted by the applicant are of sufficient quality to allow a thorough review of the data. I am able to reproduce the analyses of the primary and key secondary efficacy endpoints for each clinical study submitted.

3.2 EVALUATION OF EFFICACY

The core registration phase 3 program consists of two replicate 24-week trials, both entitled "A Phase 3, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the

Efficacy and Safety of Lumacaftor in Combination with Ivacaftor in Subjects Aged 12 Years and Older with Cystic Fibrosis, Homozygous for the *F508del-CFTR* Mutation". The two studies, 809-103 and 809-104, had the same design each with 3 arms: LUM 600mg qd/IVA 250mg q12h, LUM 400mg/IVA 250mg q12h, or placebo.

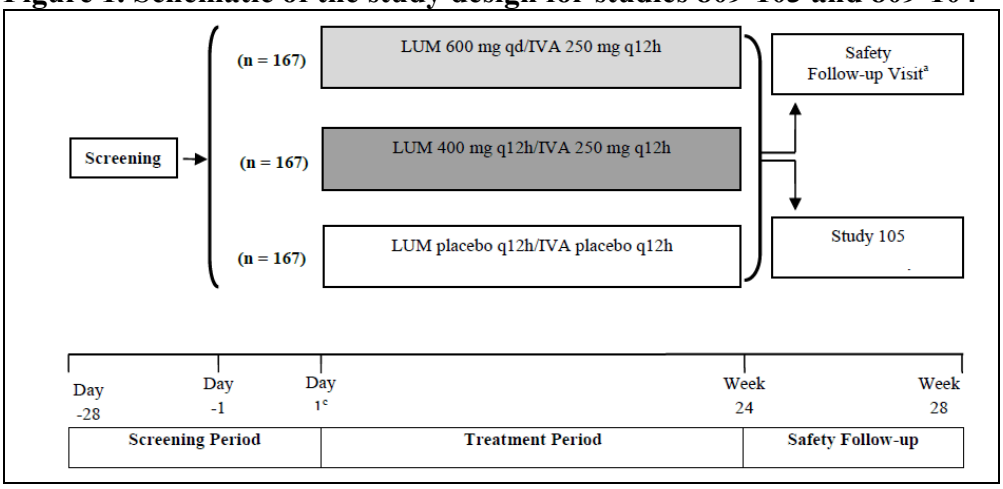
According to the applicant, an ivacaftor monotherapy arm was not included in the phase 3 studies because the results from study 770-104 had not demonstrated a clinically meaningful benefit after 16 weeks of treatment with ivacaftor monotherapy (150 mg q12h) in subjects homozygous for the *F508del-CFTR* mutation. This is reflected in the currently approved label under limitation of use, "Not effective in patients with CF who are homozygous for the *F508del* mutation in the *CFTR* gene." The applicant also noted that a lumacaftor monotherapy arm was not evaluated because of the poor efficacy of lumacaftor monotherapy demonstrated in phase 2 clinical studies, coupled with a low response in vitro to lumacaftor alone in airway epithelial cells from patients homozygous for the *F508del-CFTR* mutation. The division agreed with applicant during the IND phase that ivacaftor and lumacaftor as monotherapies was not necessary in phase 3 studies. Due to lack of monotherapy arms in the phase 3 trials, it is difficult if not impossible to evaluate the contribution of each individual component to the combination therapy. Refer to review by David Petullo for a detailed discussion in this regard.

3.2.1 Study Design and Endpoints

Studies 809-103 and 809-104 were designed to evaluate lumacaftor in combination with ivacaftor in subjects 12 years of age and older with CF who are homozygous for the *F508del-CFTR* mutation. These were 24-week, multicenter, randomized, double-blind studies with 3 parallel groups. Enrollment was limited to subjects who were 12 years of age and older, had 40% to 90% $ppFEV_1$ at screening, and were clinically stable at the start of the study. Eligible subjects were stratified by age (<18 versus ≥18 years old), sex, and $ppFEV_1$ severity at screening (<70% or ≥70%) and randomized equally to one of 3 treatment arms: LUM 600mg qd/IVA 250mg q12h, LUM 400mg/IVA 250mg q12h, or placebo.

During the 24-week treatment period, subjects took study drug orally within 30 minutes of consumption of fat-containing food. Study drug was administered in addition to the subject's usual prescribed CF therapy. The scheduled visits consisted of baseline, randomization (Day 1), Day 3, Weeks 2, 4, 8, 12, 16, 20, and 24. At the Week 24 visit, subjects who completed all visits in the treatment period were offered the opportunity to enroll in the rollover study (VX12-809-105), which included both a double-blind treatment cohort (active study drug administered) and an observational cohort (no study drug administered). For those subjects that did not enroll in the extension study, a follow-up visit was scheduled 4 weeks after the Week 24 visit. These studies were conducted in over 90 clinical centers in North America, Europe, and Australia. The timeline for these studies is shown in Figure 1.

Figure 1. Schematic of the study design for studies 809-103 and 809-104



Source: Clinical Overview, Figure 1

In both studies, the primary endpoint was absolute change from baseline in $ppFEV_1$ at Week 24, assessed as the average of the treatment effects at Week 16 and at Week 24. The baseline value was defined as the most recent non-missing measurement (scheduled or unscheduled) before the initial administration of study drug. Absolute change from baseline was calculated as post-baseline value minus baseline value.

The protocol defined several secondary endpoints, including five key secondary efficacy endpoints summarized below:

1) Relative change from baseline in $ppFEV_1$ at Week 24, assessed as the average of the treatment effects at Week 16 and at Week 24
2) Absolute change from baseline in body mass index (BMI) at Week 24
3) Absolute change from baseline in the Cystic Fibrosis Questionnaire-Revised (CFQ-R) respiratory domain at Week 24.
4) Response defined as ≥5% increase in relative change from baseline in $ppFEV_1$ at Week 24 (average of the treatment effects at Week 16 and at Week 24)
5) Number of pulmonary exacerbations through Week 24.

Relative change from baseline was calculated in as: $100 \times \frac{post\text{-}baseline\ value - baseline\ value}{baseline\ value}$.

Additional secondary endpoints included the time-to-first pulmonary exacerbation, the incidence of having at least 1 pulmonary exacerbation, and number of days with pulmonary exacerbations.

3.2.2 Statistical Methodologies

Each study planned to enroll 501 subjects (167 subjects for each treatment group) which were based on the protocol-defined efficacy endpoint of absolute change from baseline in $ppFEV_1$ at Week 24, with the following assumptions:

- A treatment difference of mean absolute change from baseline in $ppFEV_1$ of 5% between the active and placebo treatment groups, and a common standard deviation (SD) of 8%
- A 10% missing data/drop-out rate
- A 2-sided, 2-group, t-test of equal means
- An alpha of 0.025 to address the multiplicity across the 2 active doses and ensure an overall Type I error of 0.05

These studies had approximately 99% power to detect a treatment difference of 5% in absolute change of $ppFEV_1$ between either dose of LUM/IVA and placebo. As noted in Section 2.1.2, an observed small effect could be statistically significant but clinically difficult to interpret.

For the primary efficacy endpoint, absolute change from baseline in $ppFEV_1$ at Week 24, the primary analysis was to test the difference between each active combination treatment group versus placebo using a mixed model with repeated measures (MMRM). Both on-treatment measurements and measurements after treatment discontinuation (for subjects who discontinued dosing early) were included in primary analyses. The MMRM analysis included subject as a random effect, treatment, visit, and treatment-by-visit interaction as fixed effects, with adjustment for sex, age group at baseline, and $ppFEV_1$ severity at screening. An unstructured covariance structure was assumed to model the within-subject errors. A Kenward-Roger approximation was used for the denominator degrees of freedom. The primary result obtained from the model was the average of the treatment effects at Week 16 and at Week 24.

The analyses for the five key secondary efficacy endpoints were as follows. Relative change in $ppFEV_1$, change in BMI, and change in CFQ-R were compared using an MMRM model similar to the primary analysis. Analysis of change in BMI and change in CFQ-R also included respective baseline BMI or CFQ-R as a covariate. Response based on relative change from baseline in $ppFEV_1$ was analyzed using a 2-sided Cochran-Mantel-Haenszel (CMH) test stratified by sex, baseline age group, and $ppFEV_1$ severity at screening. Number of pulmonary exacerbations through Week 24 was based on regression analysis for a negative binomial distribution with sex, baseline age group, and $ppFEV_1$ severity at screening as covariates.

To account for the comparison of two doses of LUM/IVA to placebo, a Bonferroni correction was applied to control the overall Type I error rate at 0.05. Within each individual trial, a hierarchical testing procedure was used for the primary and key secondary endpoints at $\alpha = 0.025$ for each active treatment arm separately. The order of testing for the key secondary endpoints was as follows: Absolute change in $ppFEV_1$, relative change in $ppFEV_1$, absolute change in BMI, absolute change in CFQ-R, response rate based on improvement in $ppFEV_1$, and pulmonary exacerbations. At each step, the comparison was considered statistically significant if the p-value < 0.025 and all previous tests also met this level of significance. If a test failed, all results from subsequent tests were considered descriptive (nominal p-values).

The following analysis datasets of interest were defined in the protocol:

- All subjects Set: included all subjects in the study who were randomized or dosed.
- Full analysis set (FAS): included all subjects who received any amount of study drug. The treatment assignment for the FAS were as randomized.

- Per protocol set (PPS): included all FAS subjects without important protocol violations that might have a substantial impact on efficacy assessments. The criteria used for excluding subjects from the PPS were determined before the final data lock and were documented in the final protocol deviation plan.

The primary analysis and analysis of key secondary endpoints were performed on the FAS. The PPS was only used for supportive analyses of the primary and key secondary endpoints. There was no interim analysis during the trial.

Missing data were not imputed for efficacy analyses conducted using the MMRM approach, which made use of all available data even if a subject had missing data at some post baseline visits. Sensitivity analyses were conducted by the applicant to assess the robustness of the above analyses. For the primary and key secondary endpoints, the primary analysis was repeated with on-treatment measurements only. For the primary efficacy endpoint, a second sensitivity analysis was performed using analysis of covariance (ANCOVA) with multiple imputations. For the key secondary endpoint of number of pulmonary exacerbations, a second sensitivity analysis was performed using a stratified Wilcoxon rank-sum test.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

A total of 1122 subjects were enrolled in these two studies, of which 1108 subjects received at least 1 dose of study drug and 1082 subjects completed the trial. In study 809-103, 25 (4.6%) subjects stopped medication early and 12 (2.2%) discontinued from the study prematurely. In study 809-104, 29 (5.2%) subjects terminated study drug early and 14 (2.5%) prematurely discontinued from the study. The most common reason for discontinuation from study drug treatment was adverse events, occurring in 18 (3.3%) subjects study 809-103 and 19 (3.4%) subjects in study 809-104, respectively. Patient disposition for each study is shown in Table 1.

Table 1. Patient disposition in studies 809-103 and 809-104

	Study 809-103			Study 809-104		
	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Randomized	187	185	187	187	187	189
Never dosed	3	2	5	0	2	2
Treated	184	183	182	187	185	187
Completed treatment	180 (97.8)	172 (94.0)	172 (94.5)	182 (97.3)	176 (95.1)	172 (92.0)
Discontinued treatment	4 (2.2)	11 (6.0)	10 (5.5)	5 (2.7)	9 (4.9)	15 (8.0)
Completed study	182 (98.9)	179 (97.8)	176 (96.7)	185 (98.9)	180 (97.3)	180 (96.3)
Discontinued study	2 (1.1)	4 (2.2)	6 (3.3)	2 (1.1)	5 (2.7)	7 (3.7)
Analysis Datasets						
All Subjects Set	187	185	187	187	187	189
FAS	184	183	182	187	185	187
Patients never dosed	3	2	5	0	2	2
PPS	177	179	176	182	180	181
Drug compliance <80%	3	2	2	4	4	2
Not eligible	4	2	4	1	1	4
Safety Set	184	183	182	187	185	187

Source: Modified from Table 10-1 in Clinical Study Report

Selected demographic features for all randomized and treated patients are shown for both studies in Table 2. Within each study, subject demographics and baseline characteristics were generally balanced among the three treatment groups. In both studies, the majority of subjects were White and of non-Hispanic or non-Latino ethnicity. The median age was 23 years in study 809-103 and 24 years in study 809-104. There were 158 (28.8%) subjects in study 809-103 and 132 (23.6%) subjects in study 809-104 who were less than 18 years old.

Table 2. Demographics for treated subjects

Characteristic	Study 809-103			Study 809-104		
	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Sex, n (%)						
Male	100 (54.3)	97 (53.0)	98 (53.8)	90 (48.1)	89 (48.1)	89 (47.6)
Female	84 (45.7)	86 (47.0)	84 (46.2)	97 (51.9)	96 (51.9)	98 (52.4)
Age (years)						
n	184	183	182	187	185	187
Mean	25.0	24.7	25.5	25.7	24.3	25.0
SD	10.8	9.7	10.1	10.0	8.3	9.0
Median	22.0	23.0	23.5	24.0	23.0	24.0
Minimum	12	12	12	12	12	12
Maximum	64	54	57	55	48	54
Age groups (years), n (%)						
12 to <18	53 (28.8)	53 (29.0)	52 (28.6)	43 (23.0)	43 (23.2)	46 (24.6)
≥18	131 (71.2)	130 (71.0)	130 (71.4)	144 (77.0)	142 (76.8)	141 (75.4)
Race, n (%)						
White	183 (99.5)	180 (98.4)	176 (96.7)	186 (99.5)	183 (98.9)	185 (98.9)
Black or African American	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.5)	1 (0.5)	0 (0.0)
Asian	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
American Indian or Alaska Native	0 (0.0)	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Not collected per local regulations	1 (0.5)	1 (0.5)	2 (1.1)	0 (0.0)	0 (0.0)	0 (0.0)
Other[b]	0 (0.0)	1 (0.5)	2 (1.1)	0 (0.0)	1 (0.5)	2 (1.1)
Ethnicity, n (%)						
Hispanic or Latino	8 (4.3)	5 (2.7)	6 (3.3)	5 (2.7)	7 (3.8)	3 (1.6)
Not Hispanic or Latino	175 (95.1)	177 (96.7)	174 (95.6)	181 (96.8)	175 (94.6)	184 (98.4)
Not collected per local regulations	1 (0.5)	1 (0.5)	2 (1.1)	1 (0.5)	3 (1.6)	0 (0.0)
Region, n (%)						
North America	99 (53.8)	99 (54.1)	91 (50.0)	122 (65.2)	116 (62.7)	111 (59.4)
Europe	72 (39.1)	64 (35.0)	75 (41.2)	49 (26.2)	60 (32.4)	59 (31.6)
Australia	13 (7.1)	20 (10.9)	16 (8.8)	16 (8.6)	9 (4.9)	17 (9.1)

Source: Modified from Table 10-2 in Clinical Study Reports

Baseline characteristics are shown in Table 3. Within each study, the distributions of height, weight, body mass index (BMI), $ppFEV_1$ were similar across all three treatment groups.

Table 3. Baseline characteristics for treated subjects

Characteristic	Study 809-103			Study 809-104		
	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Weight (kg)						
n	184	183	182	187	185	187
Mean	59.1	58.6	60.6	58.5	58.2	59.2
SD	11.7	11.7	12.2	13.1	12.9	12.0
Median	57	58	60	57	58	58
Minimum	35	29	31	27	30	35
Maximum	93	90	101	98	99.8	105
Height (cm)						
n	184	183	182	187	185	187
Mean	167.1	166.1	166.7	165.8	165.8	166.0
SD	9.8	10.1	10.0	10.7	10.2	10.4
Median	167	166	166	166	167	165
Minimum	139	136	136	133	137	144
Maximum	191	190	190	190	187	196
BMI (kg/m^2)						
n	184	183	182	187	185	187
Mean	21.0	21.1	21.7	21.0	21.0	21.3
SD	3.0	2.8	3.2	2.9	3.3	2.9
Median	20.8	21.0	21.2	20.9	20.7	21.1
Minimum	14.4	14.3	14.6	14.1	14.2	14.8
Maximum	32.2	28.7	29.8	29.7	35.1	31.4
$ppFEV_1$						
n	181	182	180	185	184	185
Mean	60.5	61.2	60.5	60.4	60.5	60.6
SD	13.2	13.3	14.3	14.3	13.8	14.0
Median	60.4	61.8	58.7	60.5	60.6	61.5
Minimum	34.0	31.1	34.8	33.9	34.4	31.3
Maximum	88.0	92.3	94.0	99.8	90.4	96.5
$ppFEV_1$ at Screening, n						
<70	123 (66.8)	120 (65.6)	121 (66.5)	121 (64.7)	121 (65.4)	124 (66.3)
≥70	50 (27.2)	60 (32.8)	55 (30.2)	59 (31.6)	59 (31.9)	59 (31.6)
$ppFEV_1$ at baseline, n (%)						
<40	11 (6.0)	12 (6.6)	12 (6.6)	17 (9.1)	12 (6.5)	17 (9.1)
≥40 to <70	122 (66.3)	122 (66.7)	116 (63.7)	116 (62.0)	119 (64.3)	117 (62.6)
≥70 to ≤90	48 (26.1)	47 (25.7)	51 (28.0)	49 (26.2)	51 (27.6)	49 (26.2)
>90	0 (0.0)	1 (0.5)	1 (0.5)	3 (1.6)	2 (1.1)	2 (1.1)

Source: Modified from Table 10-3 in Clinical Study Reports,

3.2.4 Results

In both studies, randomization was stratified by age (<18 versus ≥18 years old), sex, and $ppFEV_1$ severity at screening (<70% or ≥70%). During my review, I noted some discrepancies between the coding for the stratification variables and the actual value for these measurements in the clinical datasets. Information Requests were sent to the applicant on February 23, 2015 and March 10, 2015, respectively.

According to the applicant, the rate of stratification errors was small and in all cases, subjects were included in the analyses as they were stratified within the randomization. The ages for five subjects were different at screening and baseline as the subjects had birthdays prior to the start of

treatment. In study 809-103, one subject was female but was incorrectly randomized to the male stratum. This was considered as a stratification error. The applicant provided the following explanation for discrepancies in $ppFEV_1$ at screening:

- Ten subjects in study 809-103 and five subjects in study 809-104 had stratification errors involving $ppFEV_1$ severity at screening. These subjects had $ppFEV_1$ <70% at screening but were randomized to the ≥70% stratum, or vice versa.
- There were 20 subjects in study 809-103 and 16 subjects in study 809-104 that were randomized but were missing a screening $ppFEV_1$ assessment in the clinical dataset. These subjects were randomized according to their spirometry values obtained at screening which were not transferred to the clinical database. Among these subjects, 6 in study 809-103 and 5 in study 809-104 also had no $ppFEV_1$ value at the Day 1 (baseline) visit and were excluded from the applicant's primary analysis using MMRM since the dependent variable (change from baseline $ppFEV_1$) could not be determined. The other subjects, 14 in study 809-103 and 11 in study 809-104, had non-missing $ppFEV_1$ values prior to study drug administration, and the most recent of those was used as the baseline measurement in the MMRM analysis.

To evaluate the impact of the stratification errors, I conducted a set of alternative analyses by including the actual values for baseline age, sex, and $ppFEV_1$ at screening from the clinical database. Both the primary endpoint and key secondary endpoints were analyzed and results were shown under "Reviewer" in tables throughout this document.

3.2.4.1 Primary Endpoint

The primary efficacy assessment for both studies was based on the analyses of average absolute change from baseline in $ppFEV_1$ at Week 24, assessed as the average of treatment effects at Week 16 and at Week 24. Results are shown in Table 4. As pre-specified, each dose of LUM/IVA was compared to placebo using $\alpha = 0.025$. In both studies, regardless of dose, treatment with LUM/IVA demonstrated a statistically significant improvement in $ppFEV_1$. These results were consistent when the actual values for baseline age, sex, and $ppFEV_1$ at screening from the clinical database were included in the MMRM (Reviewer's analysis).

Table 4. Absolute change from baseline in ppFEV1 at Week 24*, FAS

Statistics	Study 809-103			Study 809-104		
	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Baseline						
N	181	182	180	185	184	185
Mean (SD)	60.5 (13.2)	61.2 (13.3)	60.5 (14.3)	60.4 (14.3)	60.5 (13.8)	60.6 (14.0)
Absolute Δ from baseline at Week 24*						
N	180	176	172	183	181	180
Mean (SD)	-0.6 (6.5)	3.5 (7.0)	2.1 (7.1)	-0.5 (6.6)	2.2 (7.5)	2.6 (6.7)
Applicant						
LS mean within-group change (SE)	-0.4 (0.5)	3.6 (0.5)	2.2 (0.5)	-0.2 (0.5)	2.5 (0.5)	2.9 (0.5)
LS mean difference vs placebo (95% CI)	NA	4.0 (2.6, 5.4) *P*<0.0001	2.6 (1.2, 4.0) *P*=0.0003	NA	2.6 (1.2, 4.1) *P*=0.0004	3.0 (1.6, 4.4) *P*<0.0001
Reviewer						
LS mean within-group change (SE)	-0.5 (0.5)	3.6 (0.5)	2.1 (0.5)	-0.2 (0.5)	2.3 (0.6)	2.8 (0.5)
LS mean difference vs placebo (95% CI)	NA	4.1 (2.7, 5.5) *P*<0.0001	2.7 (1.2, 4.1) *P*=0.0003	NA	2.5 (1.0, 4.0) *P*=0.0008	3.0 (1.5, 4.4) *P*<0.0001

*Assessed as the average of the treatment effects at Week 16 and at Week 24
Source: Reviewer

Table 5 presents analysis of absolute change from baseline in $ppFEV_1$ at Week 24, which are consistent with the average of the parameter at Week 16 and at Week 24.

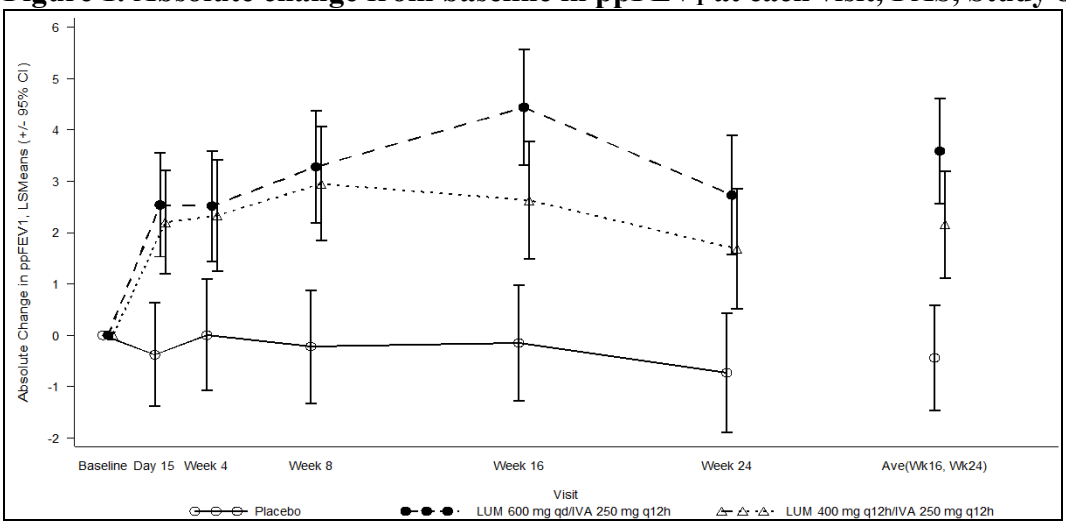
Table 5. Absolute change from baseline in ppFEV1 at Week 24, FAS

Statistics	Study 809-103			Study 809-104		
	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Baseline						
N	181	182	180	185	184	185
Mean (SD)	60.5 (13.2)	61.2 (13.3)	60.5 (14.3)	60.4 (14.3)	60.5 (13.8)	60.6 (14.0)
Absolute Δ from baseline at Week 24						
N	173	170	166	177	176	173
Mean (SD)	-0.7 (7.0)	2.7 (8.0)	1.6 (7.6)	-0.3 (7.1)	2.1 (8.2)	2.5 (7.5)
Applicant						
LS mean within-group change (SE)	-0.7 (0.6)	2.7 (0.6)	1.7 (0.6)	-0.02 (0.6)	2.3 (0.6)	2.6 (0.6)
LS mean difference vs placebo (95% CI)	NA	3.5 (1.9, 5.1) *P*<0.0001	2.4 (0.8, 4.0) *P*=0.0034	NA	2.3 (0.7, 3.9) *P* = 0.0050	2.7 (1.1, 4.2) *P*=0.0011
Reviewer						
LS mean within-group change (SE)	-0.7 (0.6)	2.7 (0.6)	1.6 (0.6)	-0.06 (0.6)	2.2 (0.6)	2.5 (0.6)
LS mean difference vs placebo (95% CI)	NA	3.5 (1.8, 5.1) *P*<0.0001	2.4 (0.7, 4.0) *P*=0.0046	NA	2.2 (0.6, 3.8) *P*=0.0074	2.6 (0.9, 4.2) *P*=0.0020

Source: Reviewer

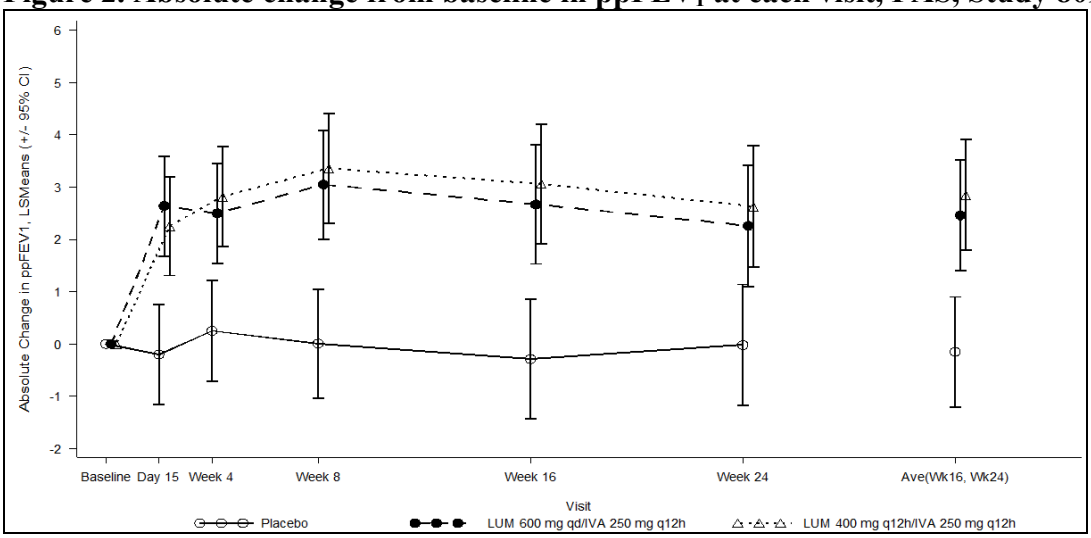
Figures 1 and 2 illustrate the absolute change from baseline in $ppFEV_1$ at each visit. For each active treatment groups in both studies, statistically significant mean absolute improvements in $ppFEV_1$ were observed at each visit when compared to the placebo group (p-values ≤ 0.05). There were no adjustments for multiplicity in these analyses. This is considered supportive of the primary analyses.

Figure 1. Absolute change from baseline in $ppFEV_1$ at each visit, FAS, Study 809-103



Source: Clinical Study Report, Figure 11-11

Figure 2. Absolute change from baseline in $ppFEV_1$ at each visit, FAS, Study 809-104

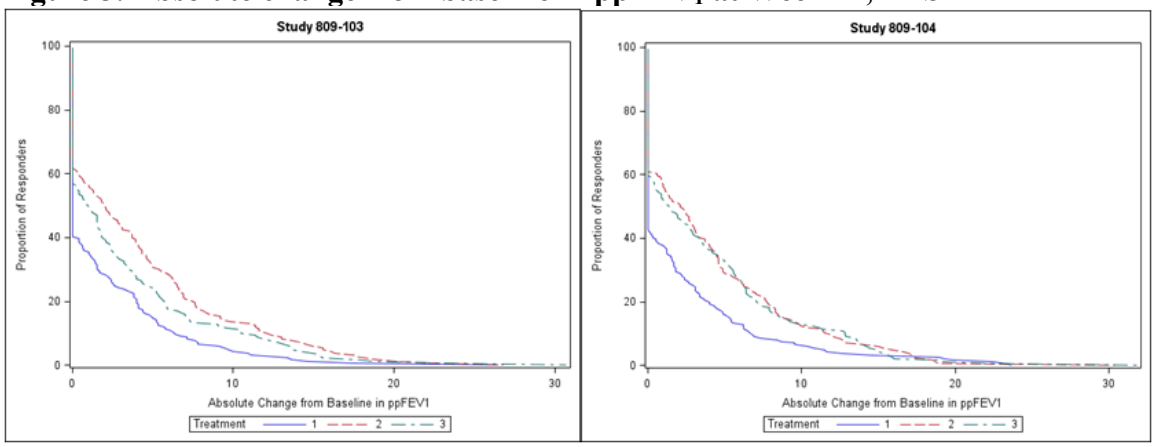


Source: Clinical Study Report, Figure 11-16

To assess the robustness of the primary endpoint analyses, the applicant performed sensitivity analyses using the MMRM approach with on-treatment measurements only and an ANCOVA model with multiple imputations for missing data, both of which generated results consistent with the primary analyses. Additional analyses based on the PPS population and a rank-based ANCOVA model also gave results that were consistent with the primary analysis.

The overall discontinuation rate was relatively low, 2.2% in study 809-103 and 2.5% in study 809-104, respectively. Nevertheless I performed a continuous responder analysis to examine the impact of missing data on the primary efficacy analysis. Patients who discontinued from the study regardless of reason were considered non-responders in this analysis. In Figure 3, the x-axis shows absolute change from baseline in $ppFEV_1$ at Week 24 and the y-axis shows the corresponding percentage of patients achieving that level of response. In both studies, at all levels of response, there were more patients treated with LUM/IVA (regardless of dose) that responded than placebo patients.

Figure 3. Absolute change from baseline in $ppFEV_1$ at Week 24, FAS



1: Placebo; 2: LUM 600 mg qd/IVA 250 mg q12h; 3: LUM 400mg/IVA 250 mg q12h
Source: Reviewer

3.2.4.2 Key Secondary Endpoints

The five key secondary endpoints were tested sequentially at α=0.025 if the primary analysis was significant. Sequential testing continued until non-significance was noted. Since the primary endpoint was significant in each study for both doses of LUM/IVA, the key secondary endpoints were tested. The results are shown in Table 6. My analysis using actual values for baseline age, sex, and $ppFEV_1$ at screening from the clinical database led to results consistent with those reported by the applicant (Table 7).

For the 1st key secondary efficacy endpoint, relative change from baseline in $ppFEV_1$, there was a significant treatment effect in favor of LUM/IVA regardless of dose in both studies. Based on the hierarchical testing procedure, the 2nd endpoint, absolute change from baseline in BMI was tested in both studies.

In study 809-103, there was an increase in BMI due to LUM/IVA treatment; however, the improvement did not reach statistical significance for either dose. Therefore, the testing hierarchy stopped at this endpoint for both active treatment groups. The results for the 3rd key secondary endpoint (CFQ-R) and 4th key secondary endpoint (Response of $\geq$5% in relative change in $ppFEV_1$) were not considered significant. Pulmonary exacerbations, the 5th key secondary endpoint, was also not considered statistically significant regardless of p-value < 0.05 as the testing hierarchy had stopped before the comparison was made.

In study 809-104, treatment with LUM/IVA resulted in significant increase in both active treatment groups for the 2nd key secondary efficacy endpoint of absolute change from baseline in BMI at Week 24. The testing continued for the 3rd key secondary efficacy endpoint, absolute change from baseline in the CFQ-R respiratory domain score at Week 24. There were improvements due to treatment with LUM/IVA but the results were not statistically significant. Based on the hierarchical testing procedure, the testing hierarchy stopped at this endpoint for both active treatment groups. The results for 4th key secondary endpoint (Response of ≥5% in relative change in $ppFEV_1$) and 5th key secondary endpoint (pulmonary exacerbations) were not considered significant and were not discussed further.

Table 6. Summary of key secondary endpoints (Applicant's analysis), FAS

Analysis	Statistics	Study 809-103			Study 809-104		
		Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
1) Relative Δ from baseline in $ppFEV_1$ at Week24* (%)	Mean	-0.3	6.4	4.0	0.0	4.4	5.3
	Difference	NA	6.7	4.3	NA	4.4	5.3
	(95% CI)		(4.3, 9.2)	(1.9, 6.8)		(1.9, 7.0)	(2.7, 7.8)
2) Absolute Δ from baseline in BMI at Week 24 (kg/m^2)	Mean	0.2	0.4	0.3	0.1	0.5	0.4
	Difference	NA	0.2	0.1	NA	0.4	0.4
	(95% CI)		(-0.0, 0.4)	(-0.1, 0.3)		(0.2, 0.6)	(0.2, 0.5)
3) Absolute Δ from baseline in CFQ-R respiratory domain score at Week 24 (points)	Mean	1.1	5.0	2.6	2.8	5.0	5.7
	Difference	NA	3.9	1.5	NA	2.2	2.9
	(95% CI)		(0.7, 7.1)	(-1.7, 4.7)		(-0.9, 5.3)	(-0.3, 6.0)
4) Response of ≥5% in relative Δ from baseline in $ppFEV_1$ at Week 24*	Yes, n(%)	41 (22.3)	85 (46.4)	67 (36.8)	42 (22.5)	85 (45.9)	77 (41.2)
	Odds ratio	NA	2.9	2.1	NA	3.0	2.4
	(95% CI)		(1.9, 4.6)	(1.3, 3.3)		(1.9, 4.6)	(1.5, 3.7)
5) Number of pulmonary exacerbations from baseline through Week 24	No. events	112	79	73	139	94	79
	Event rate/year	1.1	0.8	0.7	1.2	0.8	0.7
	Rate ratio	NA	0.7	0.7	NA	0.7	0.6
	(95% CI)		(0.5, 1.0)	(0.5, 1.0)		(0.5, 0.9)	(0.4, 0.8)

*Assessed as the average of the treatment effects at Week 16 and at Week 24

Source: Reviewer

Table 7. Summary of key secondary endpoints (Reviewer's analysis), FAS

Analysis	Statistics	Trial 809-103			Trial 809-104		
		Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
1) Relative Δ from baseline in $ppFEV_1$ at Week 24* (%)	Mean	-0.5	6.3	3.9	-0.1	4.1	5.1
	Difference	NA	6.8	4.4	NA	4.2	5.2
	(95% CI)		(4.3, 9.3)	(1.9, 7.0)		(1.6, 6.8)	(2.6, 7.8)
2) Absolute Δ from baseline in BMI at Week 24 (kg/m^2)	Mean	0.2	0.4	0.3	0.1	0.5	0.4
	Difference	NA	0.1	0.1	NA	0.4	0.4
	(95% CI)		(-0.1, 0.3)	(-0.1, 0.3)		(0.2, 0.6)	(0.2, 0.6)
3) Absolute Δ from baseline in CFQ-R respiratory domain score at Week 24 (points)	Mean	1.2	5.5	2.9	3.1	5.5	6.2
	Difference	NA	4.3	1.6	NA	2.4	3.1
	(95% CI)		(1.0,7.5)	(-1.6, 4.9)		(-0.7, 5.6)	(-0.1, 6.2)
4) Response of ≥5% in relative Δ from baseline in $ppFEV_1$ at Week 24*	Yes, n(%)	41 (22.3)	85 (46.4)	67 (36.8)	42 (22.5)	85 (45.9)	77 (41.2)
	Odds ratio	NA	2.9	2.1		2.8	2.3
	(95% CI)		(1.9,4.6)	(1.3, 3.4)	NA	(1.8, 4.4)	(1.5, 3.6)
5) Number of pulmonary exacerbations from baseline through Week 24	No. events	112	79	73	139	94	79
	Event rate/year	1.1	0.8	0.7	1.2	0.8	0.7
	Rate ratio	NA	0.7	0.7	NA	0.7	0.6
	(95% CI)		(0.5, 1.0)	(0.5, 0.9)		(0.5, 0.9)	(0.4, 0.8)

*Assessed as the average of the treatment effects at Week 16 and at Week 24
Source: Reviewer

3.3 EVALUATION OF SAFETY

The safety information for LUM/IVA was derived primarily from the two phase 3 studies (809-103 and 809-104) which included a total of 1108 patients: 369 patients on LUM 600mg qd/IVA 250mg q12h, 369 patients on LUM 400mg/IVA 250mg q12h, and 370 patients on placebo.

As reported by the applicant, there were no deaths in either study. Serious adverse events (SAE) occurred more commonly in placebo patients compared to LUM/IVA patients. Adverse events (AE) leading to treatment discontinuation were more common in LUM/IVA groups compared to placebo. Liver-related SAEs and AEs leading to discontinuation, while not common, occurred in LUM/IVA groups, but not in placebo. Respiratory symptom related AEs occurred sooner after dosing and more commonly in LUM/IVA patients compared to placebo. Additionally, respiratory symptom related SAEs and AEs leading to discontinuation, while rare, occurred in LUM/IVA patients, but not in placebo patients. With regard to effects on menstruation, adverse events related to menstrual abnormalities were more common in women in the LUM/IVA groups compared to placebo, especially in patients on hormonal contraception. While the general analysis of deaths and adverse events did not reveal specific safety concerns, the safety data

suggest that LUM/IVA exposure might be associated with liver, respiratory, and menstrual related adverse events.

Please refer to the review by Medical Officer, Dr. Robert Lim, for discussion of safety evaluation.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The applicant conducted subgroup analyses for the primary efficacy endpoint to assess the consistency of treatment effects across demographic and clinical subgroups including gender, age, region, and $ppFEV_1$ severity at screening. The treatment effects were evaluated in each subgroup using the same MMRM model as used for the primary analysis. Since these were descriptive analyses, overall type I error was not protected. Results from each individual study are presented in this section.

The conclusions were consistent with those from the study population as a whole. For each subgroup, analysis of average absolute change from baseline in $ppFEV_1$ favored LUM/IVA regardless of dose. For some subgroups, interpretation of outcomes should be treated with caution due to the small number of subjects.

4.1 Gender, Race, and Age

The average absolute change from baseline in $ppFEV_1$ is summarized according to gender and age categories (Table 8). Since the majority of these patients were white (98.2% to 99.1%) and of non-Hispanic or non-Latino ethnicity (95.8% to 96.6%), a subgroup analyses for race was not performed. Approximately 28.8% of patients in study 809-103 and 23.6% of patients in study 809-104 were less than 18 years old. Regardless of dose, there was a treatment benefit for LUM/IVA across all gender and age subgroups.

Table 8. Absolute change from baseline in ppFEV1 at Week 24* by subgroup, FAS

	Statistics	Trial 809-103			Trial 809-104		
		Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Gender	**Male,** n	96	93	94	87	87	84
	Average Δ from baseline (SE)	-0.5 (0.7)	3.1 (0.7)	2.1 (0.7)	-0.5 (0.8)	2.6 (0.8)	3.2 (0.8)
	Difference from placebo (95% CI)	NA	3.7 (1.7, 5.6)	2.6 (0.7, 4.6)	NA	3.1 (0.9, 5.3)	3.8 (1.5, 6.0)
	Female, n	84	83	78	96	94	96
	Average Δ from baseline (SE)	-0.3 (0.8)	4.2 (0.8)	2.3 (0.8)	0.1 (0.7)	2.3 (0.7)	2.5 (0.7)
	Difference from placebo (95% CI)	NA	4.5 (2.4, 6.5)	2.6 (0.6, 4.7)	NA	2.2 (0.3, 4.1)	2.3 (0.4, 4.2)
Age	**≥12 to<18 years,** n	49	51	49	42	42	44
	Average Δ from baseline (SE)	0.5 (1.2)	4.8 (1.2)	3.7 (1.2)	0.8 (1.3)	2.7 (1.3)	2.4 (1.3)
	Difference from placebo (95% CI)	NA	5.2 (1.9, 8.6)	4.1 (0.8, 7.5)	NA	2.0 (-1.7, 5.6)	1.7 (-2.0, 5.3)
	≥18 years, n	131	125	123	141	139	136
	Average Δ from baseline (SE)	-0.6 (0.5)	3.0 (0.6)	1.4 (0.6)	-0.7 (0.6)	2.1 (0.6)	2.8 (0.6)
	Difference from placebo (95% CI)	NA	3.6 (2.1, 5.1)	2.0 (0.6, 3.5)	NA	2.8 (1.3, 4.4)	3.5 (1.9, 5.0)

*Assessed as the average of the treatment effects at Week 16 and at Week 24
Source: Clinical Study Report

4.2 Other Special/Subgroup Population

As studies 809-103 and 809-104 were conducted in over 90 centers worldwide, the applicant performed subgroup analyses on the primary endpoint by region. I also conducted an analyses based on lung function since approximately one-third of patients had ppFEV$_1$ ≥70% at screening. Table 9 presents subgroup analyses by region and lung function at screening.

Regardless of region or lung function at screening, when the primary endpoint of absolute change from baseline in ppFEV$_1$ was considered, treatment with LUM/IVA demonstrated a treatment benefit compared to placebo.

Table 9 Absolute change from baseline in ppFEV1 at Week 24* by subgroup, FAS

	Statistics	Trial 809-103			Trial 809-104		
		Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250	Placebo	LUM 600 /IVA 250	LUM 400 /IVA 250
Region	**North America,** n	99	95	87	120	116	108
	Average Δ from baseline (SE)	0.0 (0.7)	3.4 (0.7)	1.8 (0.8)	-0.7 (0.7)	2.4 (0.7)	2.9 (0.7)
	Difference from placebo	NA	3.4	1.8	NA	3.1	3.6
	(95% CI)		(1.5, 5.4)	(-0.2, 3.7)		(1.3, 5.0)	(1.8, 5.5)
	Europe, n	68	62	69	47	56	55
	Average Δ from baseline (SE)	-1.4 (0.9)	3.7 (0.9)	3.0 (0.9)	0.5 (1.0)	1.6 (0.9)	2.5 (1.0)
	Difference from placebo	NA	5.1	4.3	NA	1.1	2.1
	(95% CI)		(2.6, 7.5)	(2.0, 6.70)		(-1.6, 3.8)	(-0.6, 4.7)
	Australia, n	13	19	16	16	9	17
	Average Δ from baseline (SE)	0.4 (1.8)	4.3 (1.3)	0.7 (1.5)	1.3 (1.7)	7.9 (2.3)	3.7 (1.7)
	Difference from placebo	NA	3.8	0.3	NA	6.6	2.4
	(95% CI)		(-0.5, 8.1)	(-4.1, 4.7)		(1.0, 12.3)	(-2.5, 7.2)
ppFEV$_1$ at Screening	**<70%,** n	123	115	117	121	118	122
	Average Δ from baseline (SE)	-0.1 (0.6)	3.4 (0.6)	2.9 (0.6)	-0.9 (0.7)	2.1 (0.7)	2.7 (0.7)
	Difference from placebo	NA	3.4	3.0	NA	3.1	3.6
	(95% CI)		(1.8, 5.1)	(1.3, 4.6)		(1.4, 4.8)	(1.9, 5.2)
	≥70%, n	49	59	52	57	59	56
	Average Δ from baseline (SE)	-1.0 (1.1)	4.5 (1.0)	1.2 (1.1)	1.1 (1.0)	2.4 (1.0)	2.7 (1.0)
	Difference from placebo	NA	5.5	2.2	NA	1.4	1.6
	(95% CI)		(2.5, 8.4)	(-0.8, 5.2)		(-1.5, 4.2)	(-1.3, 4.5)

*Assessed as the average of the treatment effects at Week 16 and at Week 24
Source: Clinical Study Report

5 SUMMARY AND CONCLUSIONS

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE

This submission contains two replicate phase 3 studies (809-103 and 809-104) which evaluated the efficacy and safety of LUM/IVA combination in cystic fibrosis patients with homozygous *F508del CFTR* gene. Because of two LUM/IVA doses and multiple endpoints in each individual trial, a hierarchical testing procedure with Bonferroni correction at $\alpha = 0.025$ was applied. This approach adequately controlled the overall Type I error at $\alpha = 0.05$.

Results from study 809-103 and study 809-104 were very similar; both dosing regimens of LUM/IVA demonstrated superiority over placebo in terms of spirometry function. In study 809-103, treatment with LUM 600 mg qd/IVA 250 mg q12h and LUM 400 mg /IVA 250 mg q12h resulted in statistically significant improvements in ppFEV$_1$ over placebo of 4.0% and 2.6%, respectively. In study 809-104, the average treatment effect compared to placebo was 2.6% for LUM 600mg qd/IVA 250mg q12h and 3.0% for LUM 400mg /IVA 250mg q12h, respectively. The results were similar no matter how absolute change from baseline in ppFEV$_1$ was determined, either at Week 24 or as the average of the treatment effects at Week 16 and at Week 24. The findings were also consistent regardless of age, sex, geographic region, disease severity at screening. Missing data was minimal and was not a concern.

Even though improvements from baseline were observed for the five key secondary endpoints for both doses of LUM/IVA, only the relative change from baseline in ppFEV$_1$ provided

replicated evidence of a treatment benefit. The testing hierarchy stopped at the analysis of BMI for study 809-103 and at the analysis of CFQ-R for study 809-104. Results from all subsequent analyses, including change in CFQ-R respiratory domain, response rate based on improvement in $ppFEV_1$, and pulmonary exacerbations were not considered statistically significant regardless of p-values.

While statistically significant improvements over placebo in $ppFEV_1$ were identified in both trials, the effect sizes of 2.6% to 3.0% for the claimed dose (LUM 400 mg /IVA 250 mg q12h) were considered small with unclear clinical relevance. The issue was particularly important in the context of previous results from the ivacaftor development program where a mean absolute change from baseline in$ppFEV_1$, compared to placebo of 1.7% (95% CI: [-0.6%, +4.1%]; P=0.15) was considered as not having a clinically meaningful treatment effect. The observed statistical significance in the current phase 3 trials could be due to large sample size. With about 187 subjects enrolled in each group, the study had the ability to detect a mean difference in absolute change in $ppFEV_1$ as small as 1.65%.

It should be noted that the treatment effects discussed in this review are in comparison to placebo, whereas combination products typically need to demonstrate a significant difference over each mono component. The two phase 3 studies by designed did not include a lumacaftor or ivacaftor monotherapy arm, thus did not allow the contribution of each individual component to be evaluated. Refer to the statistical review by David Petullo for additional discussion on this issue.

5.2 CONCLUSIONS AND RECOMMENDATIONS

Overall, the assessment of efficacy in the phase 3 clinical trials demonstrated that LUM/IVA provided consistent statistically significant benefits over placebo in terms of changes in $ppFEV_1$. However, the clinical meaningfulness of the magnitude of the improvement remains to be understood and the contribution of lumacaftor and ivacaftor to the combination product cannot be evaluated in the phase 3 studies.

6 LABEL REVIEW

The focus of the labeling review will be on Sections 14. Edits to the label are pending. Based on the preliminary review of the proposed label, we have the following general comments for consideration on (b) (4)

(b) (4)

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

/s/

LAN ZENG
06/01/2015

DAVID M PETULLO
06/01/2015
I concur.

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 206-038 **Applicant: Vertex** **Stamp Date: 11/05/2014**

Drug Name: Orkambi® **NDA/BLA Type: NDA**

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

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	X			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? __Yes_____

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

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

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			X	
Appropriate references for novel statistical methodology (if present) are included.	X			
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	X			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			

Studies to be reviewed: Trial VX12-809-103, Trial VX12-809-104

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

Reviewing Statistician	Date
Supervisor/Team Leader	Date

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

/s/

LAN ZENG
12/30/2014

DAVID M PETULLO
12/31/2014

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

Statistical Review and Evaluation

CARCINOGENICITY STUDIES

IND/NDA Number:	NDA 20-6038
Drug Name:	VX-809
Applicant:	Sponosor: Vertex Pharmaceuticals, Inc 50 Northern Ave. Boston, MA 02210
	Test Facility: (b) (4)
Documents Reviewed:	Electronic data submitted on September 15, 2014.
Review Priority:	Standard
Biometrics Division:	Division of Biometrics -6
Statistical Reviewer:	Min Min, Ph.D.
Concurring Reviewer:	Karl Lin, Ph.D.
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Reviewing Pharmacologist:	Andrew Goodwin Ph.D.
Project Manager:	
Keywords:	Carcinogenicity, Dose response

Table of Contents

1. Background

In this submission the sponsor included reports of an animal carcinogenicity study in transgenic mice. These studies were intended to further assess the carcinogenic potential of VX-809 in hemizygous Tg.rasH2 mice when administrated by oral gavage at appropriate drug levels for a period of 26 weeks. Results of this review have been discussed with the reviewing pharmacologist Dr. Goodwin.

2. Study design

Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were three treated groups, one water control and one vehicle control group. One hundred and twenty five Tg.rasH2 transgenic mice of each sex were randomly allocated to treated and control groups in equal size of 25 animals. 10 male mice and 10 female mice were used in the positive control group. The dose levels of treated groups were 200, 700 and 2000 mg/kg/day for males and 200, 500 and 1500 mg/kg/day for females. In this review these dose groups would be referred to as the low, medium, and high dose group, respectively. Animals in Groups 1-7 were treated once daily by gavage for up to 26 weeks, as follows: Group 1 animals received the vehicle (0.5% methylcellulose (400 cps), 0.5% Tween 80 and 0.05% simethicone in de-ionized water), Group 2 animals received de-ionized water only, and animals in Groups 3-7 received VX-809 formulated in the vehicle. All treatments were administered at a dose volume of 10 mL/kg. The study design is detailed in the table below.

Protocol Group No.	Dose Levels (mg/kg/day)	Number of Animals Examined	
		Males	Females
1	Vehicle Control	25	25
2	Water Control	25	25
3	200	25	25
4	500	-	25
5	700	25	-
6	1500	-	25
7	2000	25	-
8	Positive Control	10	10

All animals were observed twice daily at least 6 hours apart for moribundity and mortality. In the Main cohort, animals surviving until Week 26 were sacrificed by CO2 overdose and necropsied. Prior to sacrifice, animals were weighed to the nearest 0.1 gram. A complete necropsy was performed for the terminal sacrificed animals and for any moribund sacrificed animals or animals found dead.

The following tissues/organs were collected from all Main cohort animals.

Tissues/Organs	
Adrenal glands	Nasal cavity
Aorta	Ovaries
Bone (femur and sternum)	Pancreas
Bone marrow (femur and sternum)	Parathyroid glands
Brain	Pituitary gland
Epididymides	Prostate gland
Esophagus	Salivary gland
Eyes	Sciatic nerve
Gall bladder	Seminal vesicles
Gross lesions	Skeletal muscle (thigh)
Harderian gland	Small intestine (duodenum, jejunum, and ileum)
Heart	
Kidneys	Spinal cord (cervical, thoracic, and lumbar)
Large intestine (cecum, colon, rectum)	
Liver	Spleen
Lungs and bronchi	Stomach
Lymph nodes (mesenteric and mandibular)	Testes
	Thymus
Skin from mammary area (male and female mice)	Thyroid glands
	Trachea
Mammary gland (females only)	Urinary bladder
	Uterus
	Vagina

2.1. Sponsor's analyses

2.1.1. Survival analysis

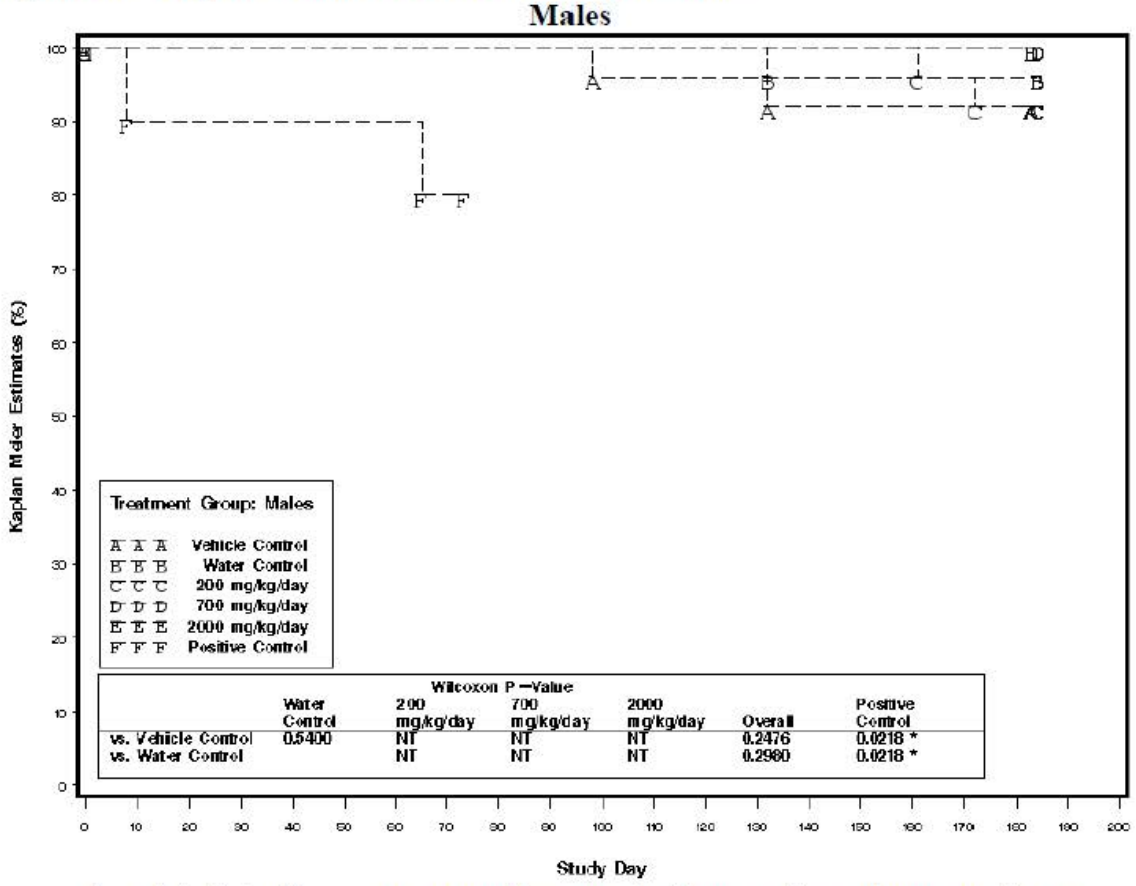
Kaplan-Meier estimates of group survival rates were calculated, by sex, and shown graphically. The generalized Wilcoxon test for survival was used to compare the homogeneity of survival rates across the control and test article groups, by sex, at the 0.05 significance level. If the survival rates were significantly different, the generalized Wilcoxon test was used to make pairwise comparisons of each test article group with each of the control groups.

The survival rate of the positive control group was compared to each of the control groups with the generalized Wilcoxon test. Survival times in which the status of the animal's death was classified as terminal sacrifice (including intermittent sacrifice) were considered censored values for the purpose of the Kaplan-Meier estimates and survival rate analyses. There were no accidental deaths in this study.
All tests were conducted at the 0.05 and 0.01 significance level without correction for multiple tests.

Sponsor's findings:

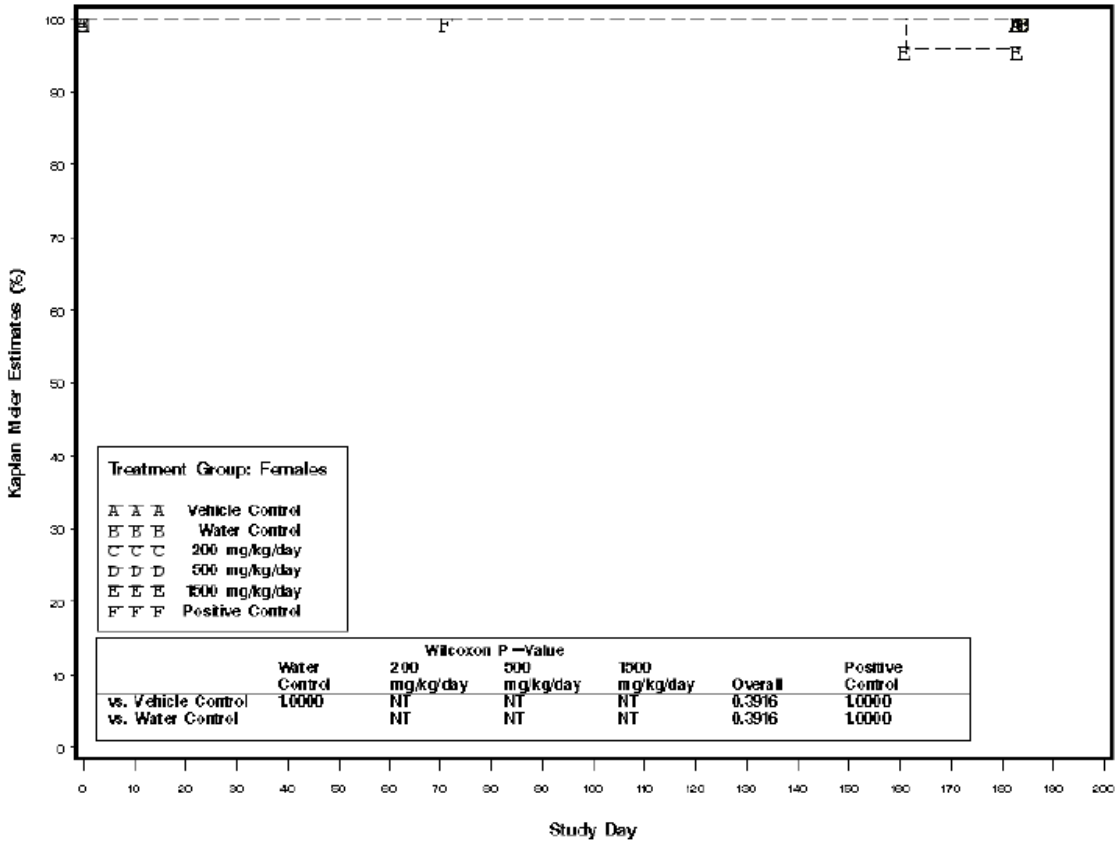
Among males, there was a statistically significant difference in survival rates when comparing the positive control to both the vehicle and water control groups. No other comparisons were statistically significant among males. There were no statistically significant survival rate findings among females.

Figure 6.3 Kaplan-Meier Estimates of Survival: Males



* - statistically significant; NT – Not tested due to non-significant overall comparison across all groups.

Figure 6.4 Kaplan-Meier Estimates of Survival: Females



* - statistically significant; NT – Not tested due to non-significant overall comparison across all groups.

2.1.2. Tumor data analysis

The incidence of tumors were analyzed, by sex, using Peto's mortality-prevalence method, without continuity correction, incorporating the context (incidental, fatal, or mortality-independent) in which tumors were observed. The following fixed intervals were used for incidental tumor analyses: days 1 - end of study (up to, but not including, scheduled terminal sacrifices), and scheduled terminal sacrifice. All tumors in the scheduled terminal sacrifice interval were considered incidental for the purpose of statistical analysis.

Tumors classified as mortality-independent were analyzed with Peto's mortality independent method incorporating the day of detection. Each diagnosed tumor type was analyzed separately. Analysis of combined tumor types and/or organs was performed. All metastases and invasive tumors were considered secondary and not included in the analyses.

A 1-sided comparison of each test article treated group with each of the control groups was performed. The water and vehicle control groups were compared with a 2-sided test. An exact permutation test was conducted for analyses with low tumor incidence.

The positive control was compared to each of the control groups with a 1-sided Fisher's exact test at both the 0.01 and 0.05 significance levels. Only the following tumors were statistically analyzed in the positive control animals: alveolar-bronchiolar adenoma, alveolar-bronchiolar carcinoma, and hemangiosarcoma in the spleen.

Tumor data was evaluated for statistical significance at both the 0.01 and 0.05 levels. All required statistics were determined using either Provantis (Tables and Statistics Version 8.4.0.1), SAS®, or Minitab® (Version 16.1.0).

Sponsor's findings:

In the test article treated groups, the incidences of single adenomas, multiple adenomas and carcinomas were comparable to those in the vehicle control (Group 1) and the DI water control (Group 2) and fell within the historical control range established at (b) (4) There was a statistically significant increase in the incidence of all pulmonary tumors in Group 5 (700 mg/kg/day) male mice when compared to the vehicle control (Group 1), but not the DI water control (Group 2). However, this statistically significant increase was not considered to be biologically or toxicologically significant, because: 1) the incidence of each of the pulmonary tumors in Group 5 fell within the historical control range, 2) the increase was significant because of zero incidence of pulmonary tumors in the vehicle control (which is at the lower end of the historical control range), 3) there were two early deaths in the vehicle control and no early deaths in Group 5, and Group 4) there was a lack of dose dependent increase in the incidence of tumors in the test article treated male mice. In the females, there were no statistically significant differences either for incidence or for trend when the control groups were compared to the test article treated groups. There was a statistically significant increase ($p<0.05$) in the incidence of lung tumors in the positive control males and females, when compared to vehicle control mice in Group 1 and the DI water control mice in Group 2.

In both sexes, the incidence of splenic hemangiosarcomas and the combined incidence of all hemangiomas and hemangiosarcomas was comparable between the vehicle control (Group 1), the DI water control (Group 2) and the test article treated groups, and fell within the historical control range established at (b) (4) There was a statistically significant increase ($p<0.05$) in the incidence of splenic hemangiosarcomas in the positive control group when compared to the vehicle control (Group 1) and the DI water control (Group 2) treated groups.

The combined incidence of harderian gland adenomas and carcinomas was significantly increased in Group 6 females. However, this increase was primarily due to zero incidence of harderian gland adenomas and carcinomas in both the vehicle control (Group 1) and the DI water control (Group 2), which is at the lowest end of the historical control range, and the incidence of harderian gland adenomas in Group 6 being at the higher end of the historical control range. Also, since the incidence of harderian gland adenomas and carcinomas in Group 6 fell within the historical control range established at (b) (4) this increase was not considered biologically or toxicologically significant.

In conclusion, treatment of Tg.rasH2 animals with VX-809 at daily oral doses of 200, 700, and 2000 mg/kg/day (in males) and 200, 500, and 1500 mg/kg/day (in females) for 26 consecutive weeks did not increase the incidence of neoplastic lesions. Therefore, VX-809 is considered to have no carcinogenic potential at the doses evaluated in the Tg.rasH2 mouse.

2.2. Reviewer's analyses

To verify sponsor's analyses and to perform the additional analysis suggested by the reviewing pharmacologist, this reviewer independently performed survival and tumor data analyses for two sets of data: one is the vehicle control with three treated groups and the other is the water control with three treated groups. Data used in this reviewer's analyses were provided by the sponsor electronically.

2.2.1. Survival analysis

The survival distributions of animals in all five treatment groups (the water control group, the vehicle control group and three treated group) were estimated by the Kaplan-Meier product limit method. Here the positive control group is excluded. The dose response relationship and homogeneity of survival distributions were tested using the Cox test (Cox, 1972) for two sets of data. The intercurrent mortality data are given in Tables 1A and 1B in the appendix for males and females, respectively. The Kaplan-Meier curves for survival rate are given in Figures 1A and 1B in the appendix for males and females, respectively. Results for the tests for dose response relationship and homogeneity of survivals, are given in Tables 2A1, 2A2, 2B1 and 2B2 in the appendix for males and females, respectively.

Reviewer's findings: The test results showed no statistically significant dose-response relationship and statistically significant difference in mortality in either sex when compared with the vehicle control group or water control group, respectively.

2.2.2. Tumor data analysis

The tumor data were analyzed for dose response relationships and pair-wise comparisons of control group with each of the treated groups were performed using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993). One critical point for Poly-k test is the choice of the appropriate value of k. For long term 104 week standard rat and mouse studies, a value of k=3 is suggested in the literature. For short term study of 26 weeks no such suggestion is available. In this analysis the first analysis was performed using k=3. If needed, for borderline cases, the analysis was repeated with other value of k (e.g. k=2 and k=4). For the calculation of p-values the exact permutation method was used. The tumor rates and the p-values of the tested tumor types are listed in Tables 3A and 3B in the appendix for males and females, respectively.

As suggested by the reviewing pharmacologist Dr. Goodwin, this reviewer did the analysis of the combinations of adenoma+carcinoma from lung as well as hemangiosarcoma/hemangioma on a whole-animal basis in both genders and adenoma+carcinoma from harderian gland in females.

Reviewer's findings: Following tumor types showed p-values less than or equal to 0.05 either tests for dose response relationship and/or pair-wise comparisons between vehicle control or water control and each of individual treated groups.

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pair-wise Comparisons (vehicle control or water control, low, medium and high dose groups)

					2000 m				
		Vehicle	200 mg	700 mg	g				
		Cont	Low	Med	High	P_Value	P_Value	P_Value	P_Value
Organ Name	Tumor Name	N=25	N=25	N=25	N=25	Dos Resp	C vs. L	C vs. M	C vs. H
Male									
spleen	hemangiosarcoma	0	0	2	3	0.024	.	0.255	0.125
	#EXAMINED ANIMALS	(25)	(25)	(25)	(25)	.	.	.	.
	PERCENTAGE	0%	0%	8%	12%				

					1500 m				
		Vehicle	200 mg	500 mg	g				
		Cont	Low	Med	High	P_Value	P_Value	P_Value	P_Value
Organ Name	Tumor Name	N=25	N=25	N=25	N=25	Dos Resp	C vs. L	C vs. M	C vs. H
Female									
Harderian_gland	ADENOMA+CARCONOMA	0	1	1	5	0.004	0.500	0.500	0.025
	#EXAMINED ANIMALS	(25)	(25)	(25)	(25)	.	.	.	.
	PERCENTAGE	0%	4%	4%	20%				
harderian gland	adenoma	0	1	1	4	0.013	0.500	0.500	0.055
	#EXAMINED ANIMALS	(25)	(25)	(25)	(25)	.	.	.	.
	PERCENTAGE	0%	4%	4%	16%				

					1500 m				
		Water	200 mg	500 mg	g				
		Cont	Low	Med	High	P_Value	P_Value	P_Value	P_Value
Organ Name	Tumor Name	N=25	N=25	N=25	N=25	Dos Resp	C vs. L	C vs. M	C vs. H
Harderian_gland	ADENOMA+CARCONOMA	0	1	1	5	0.004	0.500	0.500	0.025
	#EXAMINED ANIMALS	(25)	(25)	(25)	(25)	.	.	.	.
	PERCENTAGE	0%	4%	4%	20%				
harderian gland	adenoma	0	1	1	4	0.013	0.500	0.500	0.055
	#EXAMINED ANIMALS	(25)	(25)	(25)	(25)	.	.	.	.
	PERCENTAGE	0%	4%	4%	16%				

. .

For vehicle control group with three treated groups:

Based on the criteria of adjustment for multiple testing of trends proposed by Lin and Rahman, the positive dose-response relationship in the incidence of hemangiosarcoma in spleen in males, adenoma and combined

adenoma and carcinoma in harderian gland were considered to be statistically significant because the p-value was less than 0.05. Also based on the criteria by Haseman, the increased tumor incidences of combined adenoma and carcinoma in harderian gland in high dose group in female mice when compared to vehicle control group was considered to be statistically significant.

For water control group with three treated groups:

Based on the criteria of adjustment for multiple testing of trends proposed by Lin and Rahman, the positive dose-response relationship in the incidence of adenoma and combined adenoma and carcinoma in harderian gland were considered to be statistically significant because the p-value was less than 0.05. Also based on the criteria by Haseman, the increased tumor incidences of combined adenoma and carcinoma in harderian gland in high dose group in female mice when compared to water control group was considered to be statistically significant.

3. Summary

In this submission the sponsor included reports of an animal carcinogenicity study in transgenic mice. These studies were intended to further assess the carcinogenic potential of VX-809 in hemizygous Tg.rasH2 mice when administrated by oral gavage at appropriate drug levels for a period of 26 weeks.

Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were three treated groups, one water control and one vehicle control group. One hundred and twenty five Tg.rasH2 transgenic mice of each sex were randomly allocated to treated and control groups in equal size of 25 animals. The dose levels of treated groups were 200, 700 and 2000 mg/kg/day for males and 200, 500 and 1500 mg/kg/day for females.

For vehicle control group with three treated groups:

Based on the criteria of adjustment for multiple testing of trends proposed by Lin and Rahman, the positive dose-response relationship in the incidence of hemangiosarcoma in spleen in males, adenoma and combined adenoma and carcinoma in harderian gland were considered to be statistically significant because the p-value was less than 0.025. Also based on the criteria by Haseman, the increased tumor incidences of combined adenoma and carcinoma in harderian gland in high dose group in female mice when compared to vehicle control group was considered to be statistically significant.

For water control group with three treated groups:

Based on the criteria of adjustment for multiple testing of trends proposed by Lin and Rahman, the positive dose-response relationship in the incidence of adenoma and combined adenoma and carcinoma in harderian gland were considered to be statistically significant because the p-value was less than 0.025. Also based on the criteria by Haseman, the increased tumor incidences of combined adenoma and carcinoma in harderian gland in high dose group in female mice when compared to water control group was considered to be statistically significant.

Min Min, Ph.D.
Mathematical Statistician

Concur: Karl Lin, Ph.D.
Team Leader, Biometrics-6

cc:
Archival NDA 206-038
Dr. Goodwin
Dr. Tiwari
Dr. Nevius
Dr. Tsong
Dr. Lin
Dr. Min

4. Appendix

Table 1A: Intercurrent Mortality Rate
Male Mice

	WATER_CONTROL		VEHICLE_CONTROL		LOW		MEDIUM		HIGH	
	NO.OF		NO.OF		NO.OF		NO.OF		NO.OF	
Week	DEATH	PERCENT	DEATH	PERCENT	DEATH	PERCENT	DEATH	PERCENT	DEATH	PERCENT
0-10	.	.	.	.	.	.	.	.	.	.
11-15	.	.	1	4%	.	.	.	.	.	.
16-20	1	4%	1	8%	.	.	.	.	.	.
21-26	.	.	.	.	2	8%	.	.	.	.
Term. Sac.	24	100.0%	23	100.0%	23	100.0%	25	100.0%	25	100.0%

Table 1B: Intercurrent Mortality Rate
Female Mice

	WATER_CONTROL		VEHICLE_CONTROL		LOW		MEDIUM		HIGH	
	NO.OF		NO.OF		NO.OF		NO.OF		NO.OF	
Week	DEATH	PERCENT	DEATH	PERCENT	DEATH	PERCENT	DEATH	PERCENT	DEATH	PERCENT
0-10	.	.	.	.	.	.	.	.	.	.
11-15	.	.	.	.	.	.	.	.	.	.
16-20	.	.	.	.	.	.	.	.	.	.
21-26	.	.	.	.	.	.	.	.	1	4%
Term. Sac.	25	100.0%	25	100.0%	25	100.0%	25	100.0%	24	100.0%

Table 2A: Intercurrent Mortality Comparison
Male Mice

Test	P-Value (across four groups)	P-Value (vehicle_control vs low)	P-Value (vehicle_control vs medium)	P-Value (vehicle_control vs high)
Dose Response	0.7477	0.9906	0.7750	0.7750
Homogeneity	0.2960	0.5965	0.3074	0.3074

Test	P-Value (across four groups)	P-Value (water_control vs low)	P-Value (water_control vs medium)	P-Value (water_control vs high)
Dose Response	0.8183	0.8898	0.8875	0.8875
Homogeneity	0.2947	0.5717	0.3173	0.3173

Table 2B: Intercurrent Mortality Comparison
Female Mice

Test	P-Value (across four groups)	P-Value (vehicle_control vs low)	P-Value (vehicle_control vs medium)	P-Value (vehicle_control vs high)
Dose Response	0.8683	1.000	1.000	0.8875
Homogeneity	0.3916	.	.	0.3173

Test	P-Value (across four groups)	P-Value (water_control vs low)	P-Value (water_control vs medium)	P-Value (water_control vs high)
Dose Response	0.8683	1.000	1.000	0.8875
Homogeneity	0.3916	.	.	0.3173

Table 3A1: Tumor Rates and P-Values for Dose Response Relationship and Pair-wise Comparisons Male Mice (vehicle control, low, medium and high dose goups)

Organ Name	Tumor Name	0 mg Cont N=25	200 mg Low N=25	700 mg Med N=25	2000 mg High N=25	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H
ALL	HEMANGIOSARCOMA+HEMA	0	1	3	3	0.069	0.510	0.125	0.125
Lung+Bronchi	ADENOMA+CARCONOMA	0	3	4	3	0.193	0.125	0.060	0.125
epididymides	hemangioma	0	0	1	0	0.253	.	0.510	.
liver	hepatocellular adeno	0	0	2	0	0.443	.	0.255	.
lungs with bron	alveolar-bronchiolar carcinoma	0	0	1	0	0.253	.	0.510	.
	adenoma	0	3	3	3	0.180	0.125	0.125	0.125
spleen	hemangiosarcoma	0	0	2	3	0.024	.	0.255	0.125
thymus	thymoma	0	0	1	0	0.253	.	0.510	.

Table 3A2: Tumor Rates and P-Values for Dose Response Relationship and Pair-wise Comparisons Male Mice (water control, low, medium and high dose goups)

Organ Name	Tumor Name	0 mg Cont N=25	200 mg Low N=25	700 mg Med N=25	2000 mg High N=25	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H
ALL	HEMANGIOSARCOMA+HEMA	1	1	3	3	0.141	0.255	0.320	0.320
Lung+Bronchi	ADENOMA+CARCONOMA	3	3	4	3	0.495	0.667	0.524	0.354
epididymides	hemangioma	0	0	1	0	0.253	.	0.510	.
liver	hepatocellular adeno	0	0	2	0	0.443	.	0.255	.
lungs with bron	alveolar-bronchiolar carcinoma	1	0	1	0	0.634	0.500	0.255	0.510
	adenoma	3	3	3	3	0.487	0.667	0.354	0.354
spleen	hemangiosarcoma	1	0	2	3	0.069	0.500	0.516	0.320
thymus	thymoma	0	0	1	0	0.253	.	0.510	.

Table 3B1: Tumor Rates and P-Values for Dose Response Relationship and Pair-wise Comparisons Female Mice (vehicle control, low, medium and high dose groups)

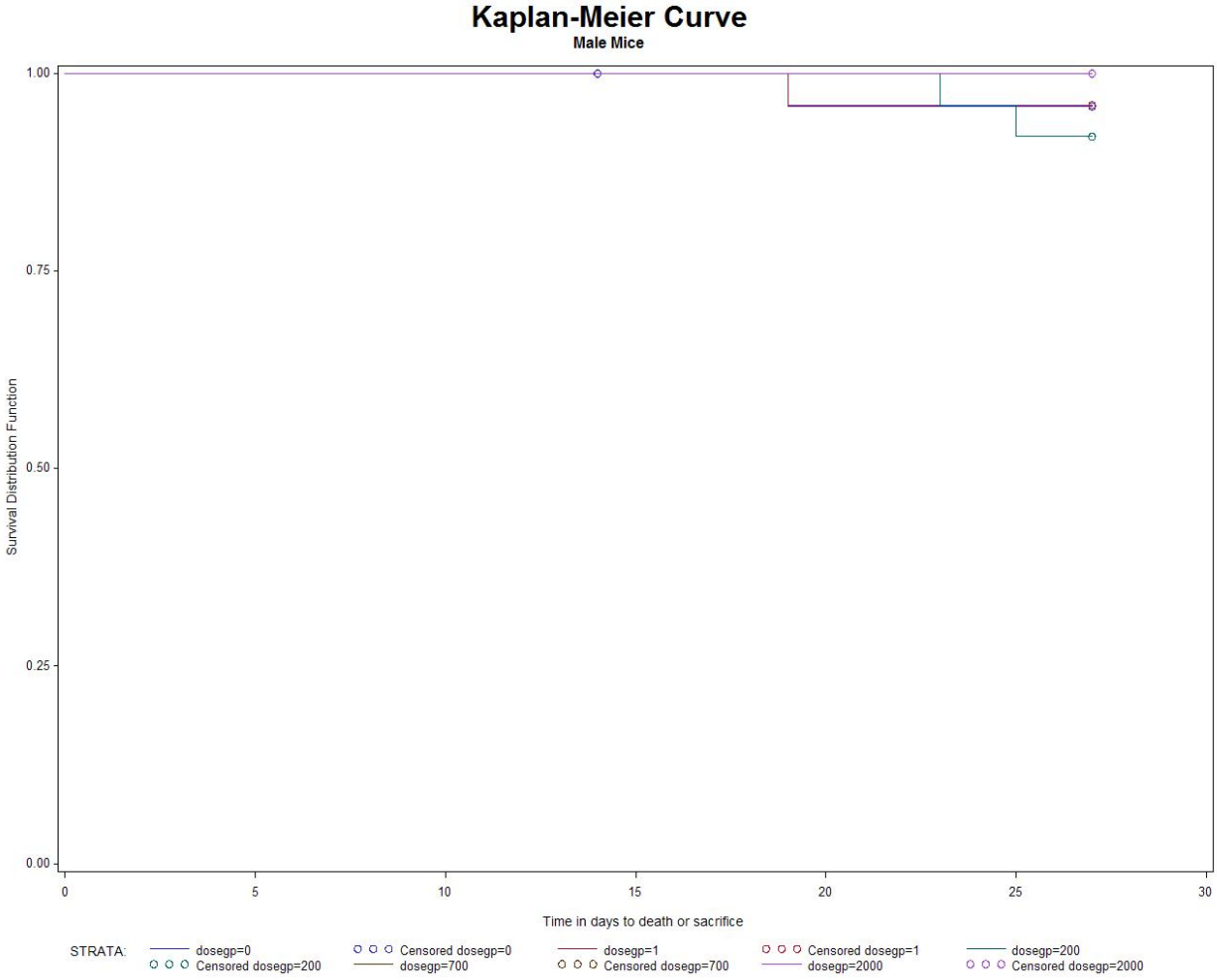
Organ Name	Tumor Name	0 mg Cont N=25	200 mg Low N=25	500 mg Med N=25	1500 mg High N=25	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H
ALL	HEMANGIOSARCOMA	0	3	1	3	0.130	0.117	0.500	0.117
Harderian_gland	ADENOMA+CARCONOMA	0	1	1	5	0.004	0.500	0.500	0.025
harderian gland	adenoma	0	1	1	4	0.013	0.500	0.500	0.055
	carcinoma	0	0	0	1	0.250	.	.	0.500
lungs with bron	alveolar-bronchiolar	4	3	2	3	0.577	0.500	0.666	0.500
spleen	hemangiosarcoma	0	2	1	1	0.418	0.245	0.500	0.500
stomach	papilloma	0	0	0	1	0.250	.	.	0.500
thymus	thymoma	0	0	0	1	0.250	.	.	0.500
vagina	hemangiosarcoma	0	1	0	1	0.313	0.500	.	0.500

Table 3B2: Tumor Rates and P-Values for Dose Response Relationship and Pair-wise Comparisons Female Mice (water control, low, medium and high dose groups)

Organ Name	Tumor Name	0 mg Cont N=25	200 mg Low N=25	500 mg Med N=25	1500 mg High N=25	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H
ALL	HEMANGIOSARCOMA	4	3	1	3	0.568	0.500	0.826	0.500
Harderian_gland	ADENOMA+CARCONOMA	0	1	1	5	0.004	0.500	0.500	0.025
harderian gland	adenoma	0	1	1	4	0.013	0.500	0.500	0.055
	carcinoma	0	0	0	1	0.250	.	.	0.500
lungs with bron	alveolar-bronchiolar	2	3	2	3	0.372	0.500	0.695	0.500
spleen	hemangiosarcoma	3	2	1	1	0.819	0.500	0.695	0.695
stomach	papilloma	0	0	0	1	0.250	.	.	0.500
thymus	thymoma	0	0	0	1	0.250	.	.	0.500
vagina	hemangiosarcoma	0	1	0	1	0.313	0.500	.	0.500

Figure 1A: Kaplan-Meier Survival Functions for Male Mice

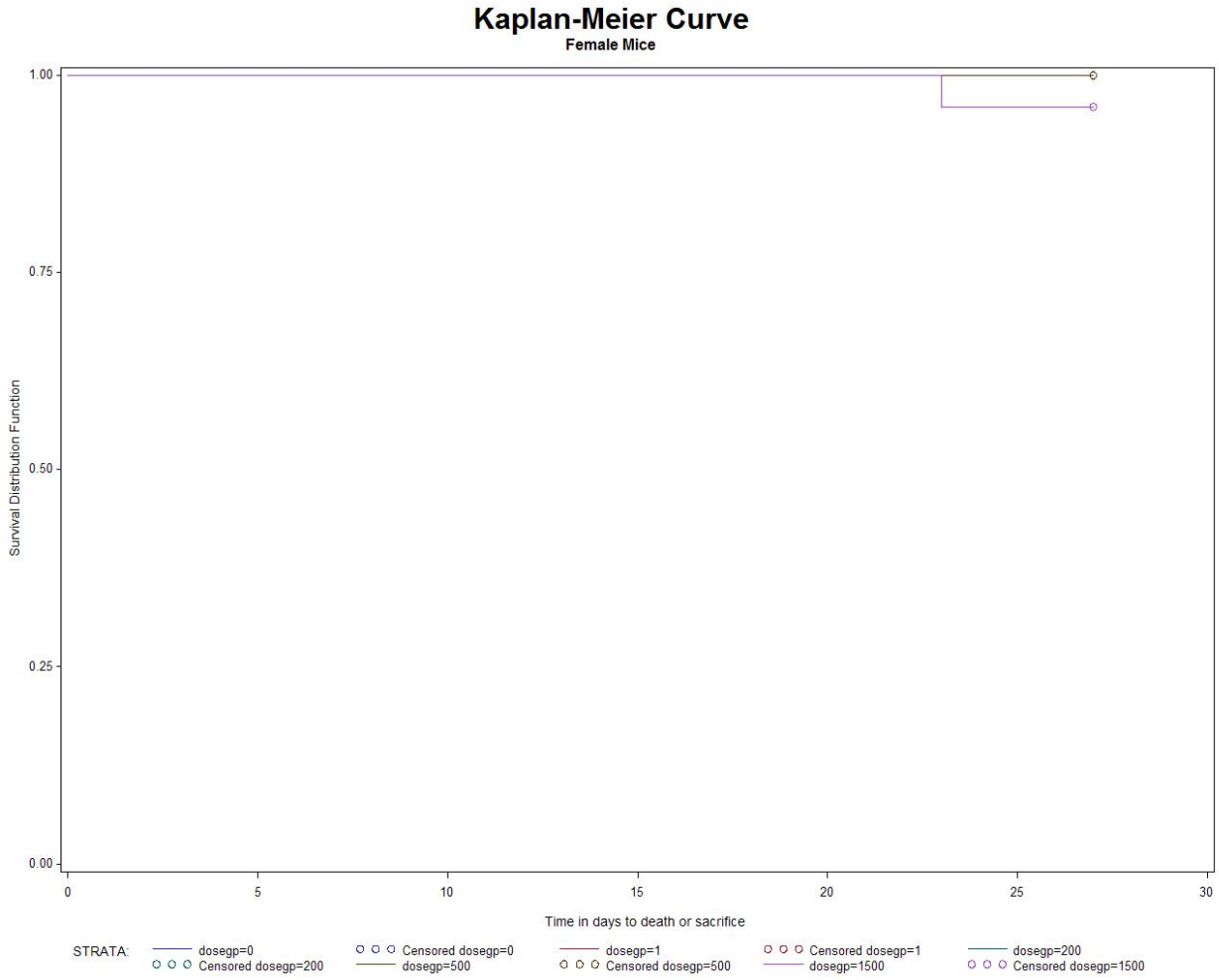
Male Mice (vehicle control, water control, low, medium and high dose groups)



X-Axis: Weeks, Y-Axis: Survival rates

Figure 1B: Kaplan-Meier Survival Functions for Female Rats

Female Mice (Vehicle control, water control, low, medium and high dose groups)



X-Axis: Weeks, Y-Axis: Survival rates

5. References:

1. Bailer AJ, Portier CJ (1988). "Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples." *Biometrics*, 44, 417-431.

2. Bieler, G. S. and Williams, R. L. (1993). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity". *Biometrics* 49, 793-801.

3. Cox D. R. (1972) "Regression models and life tables", *Journal of the Royal Statistical Society*, B, 34, 187-220.

4. Gehan (1965) "A generalized Wilcoxon test for comparing arbitrarily singly censored samples", *Biometrika*, 52, 203-223.

5. Haseman, J (1983), "A re-examination of false-positive rates for carcinogenesis studies", *Fundamental and Applied Toxicology*, 3: 334-339.

6. 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.

7. Rahman, M.A. and Lin, K.K. (2008), "A comparison of False Positive Rates of Peto and Poly-3 Methods for Long-Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman", *Journal of Biopharmaceutical Statistics,* 18:5, 849-858.

8. Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J.Wahrendorf (1980), "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva,* 311-426.

9. Tarone RE (1975), "Test for trend in life table analysis", *Biometrika*, 62: 679-82.

10. U.S. Department of Health and Human Services, "Guidance for Industry: Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals", Center for Drug E valuation and Research, Food and Drug Administration, Sliver Spring, Maryland, 2001.

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

/s/

MIN MIN
11/20/2014

KARL K LIN
11/20/2014
Concur with review