CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

207999Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA #:	NDA 207-999 OCALVIA (obeticholic acid)
Drug Name:	OCALIVA^{TM} (obeticholic acid; OCA) 5 mg tablet to be administered once daily; based on the assessment of efficacy and tolerability after 3 months, the dose may be increased to 10 mg once daily to improve response
Indication(s):	Treatment of primary biliary cholangitis (PBC) in combination with ursodeoxycholic acid (UDCA) in adults with an inadequate response to UDCA or as monotherapy in adults unable to tolerate UDCA
Applicant:	Intercept Pharmaceuticals, Inc.
Date(s):	Stamp Date: June 29, 2015
	PDUFA Goal Date: May 29, 2016
Review Priority:	New Molecular Entity (NME) 505(b)(1) with Priority Review under the PDUFA V Program (three month extension due to Major Amendment)
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Benjamin P. Vali, M.S.
Concurring Reviewers:	Yeh-Fong Chen, Ph.D.
Medical Division:	Division of Gastroenterology and Inborn Errors Products (DGIEP)
Clinical Team:	Medical Reviewer: Ruby Mehta, M.D.
	Medical Team Leader: Lara Dimick-Santos, M.D., F.A.C.S.
Project Manager:	Anissa Davis-Williams, R.N., B.S.N., M.P.H., C.P.H.M.

Keywords: NDA review, Clinical Studies

1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

There is sufficient evidence in supporting the efficacy claims for OCALIVA^{TM} (obeticholic acid; OCA), a modified bile acid and farnesoid X receptor (FXR) agonist, as a treatment for primary biliary cholangitis (PBC) in combination with ursodeoxycholic acid (UDCA) in adults with an inadequate response to UDCA or as monotherapy in adults unable to tolerate UDCA (i.e., the proposed indication). The claims reflected within the applicant's product labeling are supported by the results presented in this review.

With the intent of seeking accelerated approval, the applicant submitted the results from pivotal trial 747-301 to support the efficacy of obeticholic acid (OCA) in the treatment of PBC. Overall, the design of the 747-301 study was deemed as adequate and well-controlled from a statistical perspective, and the applicant's corresponding Statistical Analysis Plan (SAP) was adjudicated as being appropriate. There were no statistical review issues identified for this pivotal trial that would preclude product approval. The 747-301 trial results showed a significant difference in the number of responders at 12 months, pertaining to the applicant's pre-specified primary composite endpoint that utilized two hepatobiliary parameters of function and/or damage (i.e., alkaline phosphatase [ALP] and total bilirubin [TB]), between both individual OCA treatment groups and placebo. The trial results also showed that OCA substantially reduced ALP levels alone, relative to placebo, after 12 months. The currently ongoing open-label long term safety extension period suggests a sustained/durable OCA effect with respect to ALP levels. Regarding the TB levels, the changes were observed to be miniscule, and TB levels were generally stable throughout OCA treatment exposure. This may have been attributed to the enrolled 747-301 trial population primarily consisting of early stage PBC patients (as defined by the Rotterdam PBC disease staging criteria) who were being administered UDCA; these patients are understood to exhibit reasonably low and stable TB levels over time. Consequently, this may render, as questionable, any potential claim that OCA therapy maintains low and stable TB levels within this studied patient population (i.e., early stage PBC patients using UDCA) because these TB levels most likely would have stayed low and stable regardless of OCA intervention.

Although the design, statistical analyses and results of this pivotal trial appeared to be convincing and robust, the fundamental issue of this trial, and the NDA overall, was that the patients enrolled in this phase 3 study (i.e., early stage PBC patients using UDCA, as previously stated) were not adequately comparable to the patients studied by the Global PBC Group, which contained the broad spectrum of PBC disease. It was the resulting data from the Global PBC Group's research that was leveraged by the applicant to help develop the primary composite endpoint utilized in the 747-301 study. This rendered, as questionable, the overall adequacy/applicability of this trial's primary composite endpoint, which was purposed by the applicant as a basis for accelerated approval. This responder endpoint specifically incorporated 12 month changes/reductions in both ALP and TB levels assuming elevated levels for each parameter. However, the enrolled trial patients primarily exhibited only elevated ALP levels. Consequently, the statistical review team conducted an independent review using the submitted subject-level Global PBC Study data to adequately match a clinically meaningful subset of Global PBC Study patients with the aforementioned majority of enrolled patients from study 747-301, while subsequently assessing whether a 12-month reduction in ALP levels alone could be reasonably likely to predict clinical outcome (i.e., death or liver transplant) in this PBC

4

disease subpopulation. They were ultimately successful in this endeavor; please refer to Dr. Min's review for further details.

In terms of type I error control, since the 747-301 protocol only pre-specified pairwise comparisons between the individual OCA treatment groups and placebo in regards to the trial's primary composite endpoint assessed at Month 12, inferential statistics (including p-values) for only these pairwise comparison results should be presented within the final product labeling. All other 747-301 analysis results should be presented in a descriptive manner. (b) (4)

Finally, it is recommended that confirmatory clinical outcomes study 747-302 enroll the appropriate PBC patient population to ultimately verify 12-month ALP reduction alone as a surrogate reasonably likely to predict clinical benefit.

2 INTRODUCTION

2.1 Overview

On December 19, 2014, Intercept Pharmaceuticals, Inc. initiated the filing of this New Molecular Entity (NME) New Drug Application (NDA) for OCALIVA™ (obeticholic acid; OCA) in accordance with Section 505(b)(1) of the Federal Food, Drug, and Cosmetic Act and Title 21 of the Code of Federal Regulations (CFR), Part 314. The active pharmaceutical ingredient (API) of OCALIVA (5 mg tablet to be administered once daily; based on the assessment of efficacy and tolerability after 3 months, the dose may be increased to 10 mg once daily to improve response) is OCA, which is a modified bile acid and FXR agonist. Effective on January 27, 2006, the applicant had initiated clinical development of OCA, under IND 63,307, as a treatment of PBC, a hepatic condition primarily affecting females, in combination with UDCA in adults with an inadequate response to UDCA or as monotherapy in adults unable to tolerate UDCA. This proposed indication is based on the concept of developing a safe and effective therapy for the PBC patient population inadequately managed or non-responsive to existing therapies.

Primary biliary cholangitis (PBC), previously known as primary biliary cirrhosis, is an autoimmune disease which predominantly affects women (i.e., it has a purported female to male ratio of 9:1). It is a rare disease with prevalence in the United States (US) of approximately 1 per 37,000 people; the prevalence in Europe is approximately 1 per 25,000. Typically, the disease presents in middle age in patients with early stage disease; such patients do not have hepatic cirrhosis. Without treatment, the disease progresses to cirrhosis and, as the liver fails, complications develop including those related to cirrhosis, portal hypertension and hepatocellular carcinoma, the need for liver transplantation or, worst case, death. It is believed that a key result of the immunologic injury in PBC is a rise in the endogenous hepatocellular bile acids to cytotoxic levels, causing inflammatory damage which over time results in hepatic fibrosis and cirrhosis. Disease progression may take decades from diagnosis to the development of the clinical events associated with hepatic decompensation. UDCA is the only approved drug for the treatment of PBC and is used universally. However, there is a persistent medical need for better therapies in patients with an inadequate response to UDCA or those who cannot tolerate this

5

drug. Overall, PBC is considered a rare, serious, and life-threatening condition with a not sufficiently met medical need.

FXR is the nuclear receptor involved in the regulation of hepatic bile acid production and flow and the modulation of hepatic inflammation, fibrosis and regeneration. It is believed that high intracellular concentrations of bile acids are pathologically involved in cholestatic liver disease, and hence drugs that can improve bile acid flow (i.e., reducing toxic bile acid concentrations) and have anti-inflammatory effects should have beneficial effects in PBC. OCA, mechanistically as an FXR agonist, is consequently potentially such a drug whose usage can ultimately reverse PBC.

The applicant obtained *Orphan Designation* from the Office of Orphan Products Development (OOPD) on April 9, 2008. Intercept also obtained *Fast Track Designation* from the Division of Gastroenterology and Inborn Errors Products (DGIEP) on May 27, 2014. Consequently, DGIEP has agreed to receive the NDA on a rolling basis with the final component of the NDA having been submitted on June 29, 2015. A priority 8-month review cycle under the Prescription Drug User Fee Act (PDUFA) V Program has been determined for this NDA; however, a three month extension was initiated due to a Major Amendment of the review cycle that was motivated by clinical and statistical review issues.

There were a series of formal communications and meetings between the applicant and DGIEP throughout OCA's clinical development program, including all development stage-appropriate Type B meetings. The clinical aspects of the development had been the subject of many of these meetings with DGIEP. Due to the rarity of PBC and its slow natural history (as previously stated), it was very difficult to conduct clinical trials assessing long-term clinical outcomes (i.e., death or liver transplant) and hence DGIEP provided feedback regarding the possibility of pursuing accelerated approval for OCA in the treatment of PBC. Based on this advice, Intercept engaged an academic research group, i.e., the Global PBC Group, which was investigating the prognostic nature of biochemical hepatobiliary variables on transplant-free survival. This research study was named the Global PBC Study. Two of the hepatobiliary variables assessed in the Global PBC Study, i.e., ALP and TB (further details in Section 3.2.1 below) were chosen by Intercept to construct a composite endpoint for the lone pivotal trial submitted in this NDA with the hope that this composite endpoint could be used as an acceptable surrogate endpoint "reasonably likely to predict clinical benefit" to support accelerated approval as per the marketing approval regulatory pathway described in 21 CFR 314 Subpart H. The SAP for the pivotal trial analyses was designed in consult with DGIEP with a follow up meeting to finalize the plan held on October 2, 2013. The meeting also included a discussion of the data requirements for accelerated approval and next steps for the development of OCA in the treatment of PBC. Based on DGIEP guidance, it was understood by Intercept that the acceptability of the lone pivotal study to support accelerated approval is dependent on DGIEP's review of the results from this study and the determination as to whether the proposed surrogate endpoint is reasonably likely to predict clinical benefit, which is critically dependent on DGIEP's review of the Global PBC Study. Per 21 CFR 314.510, the program would also need to be supported by a confirmatory clinical outcomes study underway prior to obtaining accelerated approval. As such, two Type C meetings to discuss the design of this proposed confirmatory outcomes study were held on January 29, 2014 and July 22, 2014. Key discussion topics

6

included the proposed subject population, primary endpoint, study design and planned statistical analyses. As agreed during these meetings, Intercept addressed and incorporated DGIEP's comments, and the first version of the final confirmatory study protocol and its final SAP were submitted to IND 63,307 on October 9, 2014 (each dated October 3, 2014 and October 7, 2014, respectively). This study (i.e., trial 747-302) is currently open for enrollment in the US and Latin America, and it is in the initiation process in the European Union (EU) and other countries. There was one amendment made to the original protocol on April 29, 2015, and this version was submitted to IND 63,307 on May 6, 2015; there have been no amendments to the SAP.

The lone pivotal trial used for the basis of this accelerated approval is study 747-301 which is a phase 3, 12-month, multinational, multicenter, randomized, double-blind, placebo-controlled, parallel group study. This study utilized the aforementioned composite endpoint that incorporated ALP and TB. The double-blind/placebo-controlled portion of this trial has already been locked and unblinded. Patients completing this study had the option of additional long term open-label extension up to five years; this open-label rollover period is still currently ongoing. A total of 180 patients were targeted for enrollment, and 217 patients were ultimately recruited (with 216 actually being dosed) to participate in this study. There were two short term phase 2 studies (i.e., 747-201 and 747-202), but these studies enrolled a very small set of patients using the proposed indicated dose. In addition, safety was the important objective of these trials; hence please see the clinical review document for full details as these trials are not covered in this review.

Table 1 below presents information on the lone relevant trial contained in this submission.

Table 1
Summary Information for Relevant Clinical Trials

Type of Study; Phase	Study Identifier	Objective(s) of the Study	Study Design and Type of Control	Test Product(s); Regimen; Route	Number of Dosed Patients	Patient Diagnosis	Duration of Treatment
Safety and Efficacy; Phase 3	747-301	Efficacy, Safety, Pharmacokinetics (PK)	Multinational, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel group	OCA 5mg and 10mg; 1 daily dose; tablet orally	Total: 216	Adult subjects with a proven or likely diagnosis of PBC who are either taking UDCA or who are intolerant to UDCA	12 months with option of additional long term open-label extension up to 5 years

Source: Reviewer's Table from applicant's tabular listing of all clinical studies (eCTD Module 5 2).

2.2 Data Sources

This NDA was submitted electronically in electronic Common Technical Document (eCTD) format via the FDA Electronic Submissions Gateway (ESG). The content, including the electronic data sets and labeling information, is located in the Center for Drug Evaluation and Research (CDER) electronic document room (EDR) at the location: \\CDSESUB1\evsprod\NDA207999. Sequences 0001, 0020 and 0034 contain all the contents relevant for this review.

The clinical study report (CSR), clinical datasets and analysis datasets were reviewed for the 747-301 trial. For this study, the clinical/tabulation datasets were compliant to the CDISC/SDTM v.3.1.2 implementation guide standard. The analysis datasets for this study were compliant to the CDISC/ADaM v.1.0 implementation guide standard. Adequate data definition files (in define.xml and define.pdf formats), a reviewer's guide and software code (in .txt, format) were also submitted for this trial.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The 747-301 study utilized an electronic case report form (eCRF) within an electronic data capture (EDC) system, and the submitted data quality for this study appeared to be adequate. There were no 'Official Action Indicated' (OAI) issues from site inspections conducted by the Office of Scientific Investigations (OSI); any issue was deemed as either 'No Action Indicated' (NAI) or 'Voluntary Action Indicated' (VAI). As per OSI, the 747-301 study appeared to have been conducted adequately, and the data generated by the study appeared acceptable in support of the respective indication. It was possible to reproduce the primary analysis dataset (along with the results presented within the CSR), specifically the primary endpoint values, from the original data source. It was also possible to verify the randomized treatment assignments, and the applicant submitted documentation of data quality control/assurance procedures within Section 9.6 of the CSR.

The Statistical Analysis Plan (SAP) was finalized (the finalized version was also the initial version as well), with no subsequent amendments, on March 5, 2014. Although there were differences in some analysis approaches between this finalized SAP and what was presented within the statistical section of the finalized trial protocol, these differences were solely motivated by DGIEP advice conveyed on October 2, 2013, as stated above. There was no material presented within this finalized SAP that was not already included within the statistical section of the finalized protocol or not already advised by DGIEP. Database hard-lock for the double-blind treatment period data along with all concurrently open-label treatment data, ongoing at that time, was on January 8, 2014. The study was officially unblinded on January 15, 2014. The study data cutoff date for the ongoing long term open-label extension period was June 29, 2015 (which was the basis of the 120-Day Safety Update within the review cycle). Hence any open-label extension period data presented within this written review reflects this cutoff date. Note that any open-label extension period data accrued between January 8, 2014 and June 29, 2015 was cleaned by the clinical data management team prior to regulatory submission.

3.2 Evaluation of Efficacy

3.2.1 Background, Study Objective, Design, and Endpoints

Background

The title of the 747-301 trial is, "A Phase 3, Double Blind, Placebo Controlled Trial and Long Term Safety Extension of Obeticholic Acid in Patients with Primary Biliary Cirrhosis (PBC OCA International Study of Efficacy [POISE])". Given the paucity of the PBC population, it was understood that the 747-301 study would ultimately provide the key evidence for establishing OCA's efficacy profile within this patient population. This study was designed as a multinational (with a total of 13 participating countries within 3 geographic regions), multicenter (with a total of 59 participating study sites), randomized, double-blind, placebo-controlled, parallel group trial to evaluate the efficacy and safety of OCA in patients with PBC. The original 747-301 trial protocol was finalized on September 6, 2011, and the trial was subsequently started on March 19, 2012. The completion date of the 12-month double-blind portion of the study was on December 18, 2013; the long term open-label extension period is still currently ongoing. Three amendments (i.e., on January 18, 2012, April 4, 2012, and September 24, 2012) were made to the original protocol as pertaining to the double-blind portion of the study. Each amendment was minor and/or administrative in nature without changing key pre-specified features of the original protocol. This written review reflects the final version (i.e., dated September 24, 2012) of the double-blind portion of the protocol.

Study Objective, Design and Endpoints

The lone primary objective of this study in PBC patients was to demonstrate the efficacy of OCA, relative to placebo, based on its effects on ALP and TB. Other objectives such as assessing safety, histological, bile acid, and biomarker (i.e., not including ALP and TB) parameters were considered exploratory from a statistical perspective and hence are not presented in this review (please see the clinical review document for full details).

This phase 3 study included a screening period of up to 8 weeks, a 12-month double-blind placebo-controlled treatment period, and an open-label extension period of up to 5 years (for a total maximum participation duration of 74 months). All patients who completed or discontinued from the trial, for any reason, had a follow-up visit 4 weeks after their last dose of study medication. After the patient provided informed consent each patient underwent screening assessments to determine study eligibility. The two most significant inclusion criteria pertained to pre-treatment assessed ALP and TB values along with allowing concomitant usage of UDCA while participating in the study. Specifically, these two inclusion criteria, respectively, are as follows:

- Have <u>at least one</u> (i.e., "and/or") of the following qualifying biochemistry values
  - ALP $\geq 1.67 \times$ Upper Limit of Normal (ULN)
  - TB > ULN but < $2.0 \times$ ULN

- Taking UDCA for at least 12 months (with a stable dose for at least 3 months) prior to study start, or unable to tolerate UDCA (i.e., no UDCA usage for at least 3 months) prior to study start.

If all eligibility criteria were met, the patient was stratified into one of four groups, i.e., two factors each with two sub-categories (specified in parentheses):
- Pre-treatment ALP > 3.0×ULN and/or aspartate aminotransferase (AST) > 2.0×ULN and/or TB > ULN; ('no' for all three conditions, 'yes' to at least one of the three conditions)
- Intolerance to UDCA; ('no' hence UDCA usage for at least 12 months, with a stable dose for at least 3 months, prior to study start with the assumption of continued stable usage of UDCA throughout the study, 'yes' hence no UDCA usage for at least 3 months prior to study start with the assumption of continued non-usage of UDCA throughout the study).

The patients in each of the four possible strata were then randomized via Interactive Voice-Response System/Interactive Web-Response System (IVRS/IWRS) in a 1:1:1 ratio to receive 10 milligrams (mg) OCA, 5 mg OCA with the option to titrate up to 10 mg at Month 6 (i.e., the 'OCA Titration' treatment arm), or matching placebo. Study medication was administered orally, once daily as a single tablet, for 12 months. For all treatment arms (although specifically targeting the blinded OCA Titration treatment arm), the criteria to be eligible for up-titration at the 6 month time point/visit, assessed by the on-site investigator (and subsequently made via the IVRS/IWRS), was if the patient met any (i.e., "and/or") of the following conditions:
- ALP ≥ 1.67×ULN
- TB > ULN
- < 15% ALP reduction at Month 6 versus the mean double-blind pre-treatment ALP value(s)
- Provided adverse events (AEs) (e.g., severe pruritus) did not limit the administration of a higher dose.

Note that the visit for assessing the potential for up-titration was the Month 6A visit while the actual, if eligible, up-titration occurred at the Month 6B visit in a blinded manner. The Month 6A and 6B visits were within seven days of each other.

Following randomization, patients had five in-clinic trial visits at Week 2 and Months 3, 6, 9 and 12 to evaluate efficacy, safety, tolerability and compliance with study medication. Three central laboratories were utilized, one for each geographic region (i.e., North America, Europe, and Australia), to aid in these assessments. Patients were contacted by the trial site staff on a monthly basis between clinic visits beginning with Month 1. As previously stated, after completing the 12-month double-blind treatment period, each patient, regardless of their original randomized treatment assignment, was offered to continue on open-label OCA treatment during a long term safety extension (LTSE) period beginning at Month 12 and lasting up to 5 years. All patients participating in this LTSE period who were not being administered 10 mg OCA at the end of the double-blind treatment period would start on 5 mg OCA; patients being administered the 10 mg dose at the end of the double-blind treatment period would continue on their 10 mg dose. Clinic visits occurred every 3 months during the LTSE period and at each visit patients would be assessed to see if they qualified for 5 to 10 mg incremented up-titrations (i.e., one

10

eligible 5 to 10 mg up-titration per 3-month visit) up to a maximum trial dose of 25 mg. This maximum trial dose was later revised by an additional protocol amendment made on August 25, 2014 to not exceed 10 mg; patients who had titrated beyond a 10 mg dose prior to this amendment (i.e., on protocol versions on or before September 24, 2012) were allowed to remain on that higher dose if approved by the investigator. The criteria to be eligible for up-titrations at these visits were the same as previously presented for the Month 6 visit. All patients would continue their pre-study dose of UDCA throughout their participation in the double-blind and LTSE periods.

As previously stated in Section 3.1 above, all 747-301 trial data presented within this written review (see Section 3.2.4 below) reflect a study data cutoff date of June 29, 2015. The overall study scheme for both the double-blind and LTSE periods are shown in Figure 1 below. Note that the target sample size for the study was for 180 patients (i.e., 60 per treatment arm); a total of 217 patients were ultimately enrolled and randomized with 216 being administered at least one dose of study drug (see Section 3.2.3 below).

11

Figure 1
Schematic Diagram of Double-Blind Treatment and LTSE Periods
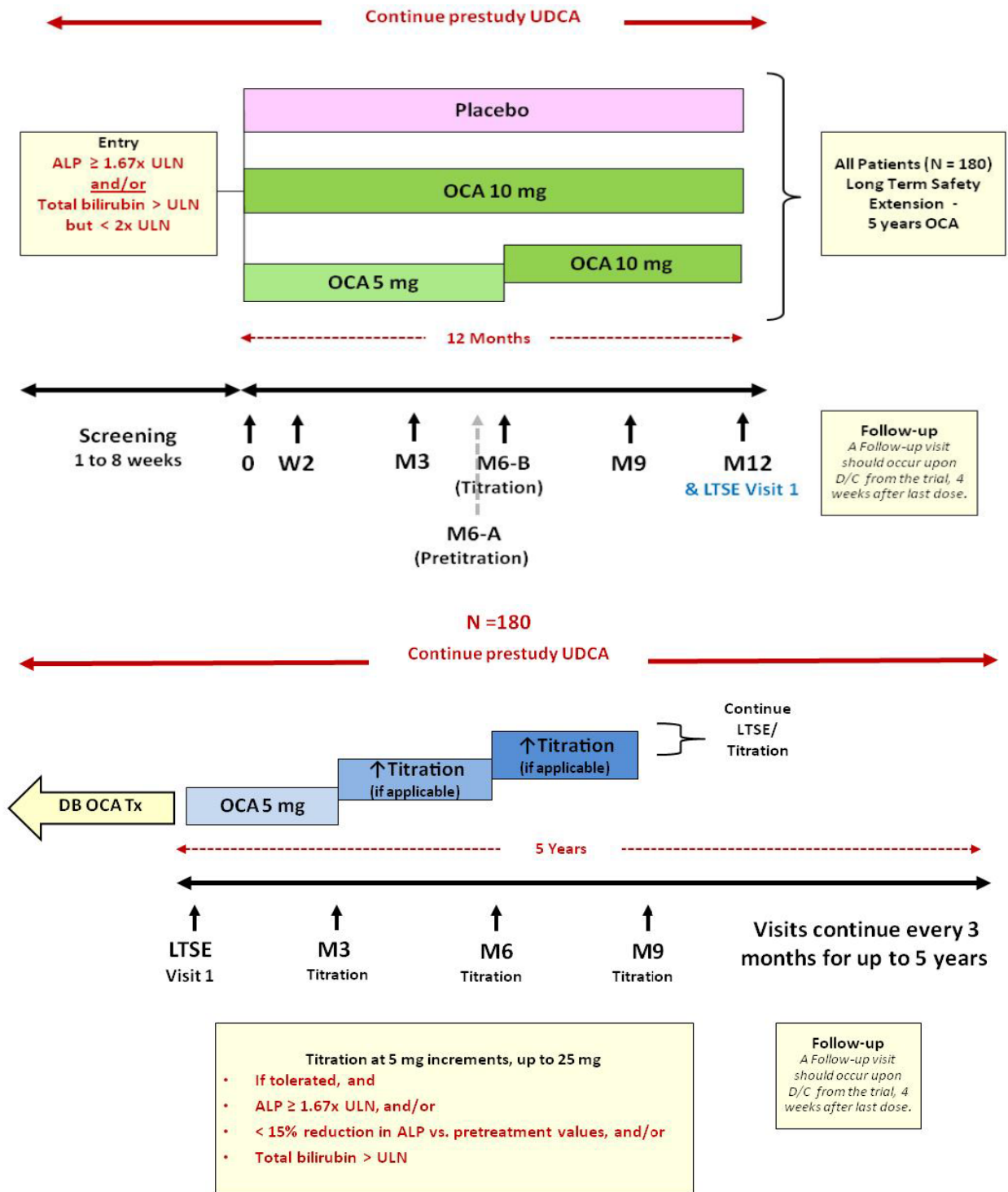
Source: Figures 1 and 2 from pgs. 31-32 of the 747-301 protocol.
Note: The August 25, 2014 protocol amendment capped the maximum total dose at 10 mg for the LTSE period.

In accordance with the study's lone primary objective, as stated above, the following primary endpoint (which was a composite endpoint) was pre-specified within the original protocol.

Primary Endpoint: ALP and TB composite response criteria at Month 12; a patient was designated as a responder if all three of the following conditions were met:
- 12-Month value of ALP < 1.67×ULN
- 12-Month value of TB ≤ ULN (i.e., within normal limits)
- ALP reduction from baseline at Month 12 ≥ 15%.

Assuming, from published literature and previous trial experience, that 40% of patients randomized to 10 mg OCA and 14% of patients randomized to placebo achieve response based on the primary composite endpoint, a sample size of 180 randomized patients (i.e., 60 patients per treatment group) would provide 90% power to detect a statistically significant difference between OCA and placebo, using a two-sided test of equality of binomial proportions at a 5% level of significance (i.e., α=0.05). It can be seen below in Section 3.2.3 that 217 patients were ultimately enrolled and randomized in this study.

Throughout the execution of this protocol, an IDMC operated according to a DMC Charter. It provided an ongoing, independent, and expert review of the safety data in order to provide risk management during the conduct of the study. Note that there were no formally planned interim analyses for this study.

Statistical Reviewer Comments:
Overall, the design of the 747-301 study was deemed adequate from a statistical perspective, and the estimated sample size was appropriate given the assumptions on the anticipated treatment effect. The only review issue pertained to the adequacy/applicability of the primary composite endpoint.

As stated above, the applicant's justification for seeking accelerated approval using this composite endpoint as a surrogate reasonably likely to predict clinical benefit is supported by the result of an academic research study conducted by a research group (founded in January 2012), i.e., the Global PBC Group, whose principle investigators are located at the Erasmus MC University Medical Center in Rotterdam, Netherlands. This research study (i.e., the Global PBC Study) was a retrospective multinational multicenter registry study that followed 4,845 adult PBC patients until they achieved clinical outcome (i.e., death or liver transplant). This registry study purported that a reduction (after the first 12 months of observation) in elevated levels of two hepatobiliary parameters of function and/or damage, i.e., ALP and TB, were reasonably likely predictors of death or liver transplant. And due to the mechanism of action of OCA, which purportedly reduces ALP and TB (see Section 3.2.4 below), the applicant subsequently leveraged the results from this study (with multiple corresponding literature publications, e.g., Lammers et. al. in 2014) to help construct this composite endpoint. The exact rationale for the applicant choosing this particular composite endpoint (i.e., the specific multiples of ULN cut points for ALP and TB along with the additional condition of a specific percent reduction in ALP) is as follows. Since 2006, there had been increased interest within the hepatology community in evaluating the relationship between biochemical markers of disease in PBC patients and the

13

development of adverse clinical outcomes. Many PBC researchers had published their own response cut point criteria as it relates to prognosticating transplant-free survival. The most publicized response criteria are listed here:

- *Barcelona (Parés 2006) – ALP reduced by at least 40% and/or ALP ≤ ULN at 1 year*
- *Global PBC Study (Lammers, 2014) – ALP ≤ 2×ULN and TB ≤ ULN at 1 year*
- *Mayo Clinic (Momah/Lindor 2012) – ALP ≤ 1.67×ULN and TB ≤ 1 mg/dl at 1 year*
- *Paris I (Corpechot 2008) – ALP ≤ 3×ULN and AST ≤ 2×ULN and TB ≤ 1 mg/dl at 1 year*
- *Paris II (Corpechot 2011) – ALP ≤ 1.5×ULN and AST ≤ 1.5×ULN and TB ≤ 1 mg/dl at 1 year*
- *Toronto II (Kumagi 2010) – ALP ≤ 1.76×ULN and/or TB ≤ ULN at 1 year*
- *Rotterdam (Kuiper 2009) – TB ≤ ULN and/or albumin ≥ lower limit of normal (LLN) at 1 year*

In addition, Kumagi in 2010 (i.e., Toronto I) also publicized the response cut point criteria of ALP ≤ 1.67×ULN alone at 2 years as being prognostic of PBC disease progression as determined by liver fibrosis on histology. Putting this all together, Intercept believed that the combination of Kumagi publications (i.e., Toronto I/II) was the most discriminating of the response criteria evaluated and consequently chose it. In addition to the combination Toronto I/II ALP and TB criteria, a minimum 15% percent reduction in ALP condition was also included to ensure that patients with a minimal effect on biochemical markers were not improperly judged as having a successful response.

It should be noted that all of these researched responder criteria/biochemical marker cut points were developed by studying a reasonably broad spectrum of PBC disease subjects (i.e., those having early, moderate, or even late stage disease). Due to the first of the two significant inclusion criteria for the 747-301 trial previously presented above, the enrollment population ended up primarily consisting of patients with screening/baseline ALP ≥ 1.67×ULN and TB ≤ ULN. It is understood within the hepatology community that this (i.e., PBC patients having TB ≤ ULN) signifies earlier stage disease patients only. The Rotterdam criteria, which is the most widely adopted PBC disease staging criteria, specifically defines early stage PBC disease as having elevated ALP (i.e., ALP > ULN), normal TB (i.e., TB ≤ ULN) and normal albumin (i.e., albumin ≥ LLN). It defines moderately advanced stage PBC disease as having elevated ALP and either abnormal TB (i.e., TB > ULN) or abnormal albumin (i.e., albumin < LLN); it defines advanced stage PBC disease as having elevated ALP, abnormal TB, and abnormal albumin. It can be seen below in Table 3 that the overwhelming majority of patients (i.e., 90.3%) enrolled in the 747-301 study were designated as early stage PBC patients by the Rotterdam criteria. (It should be noted that any reference to early stage PBC disease in this review document specifically refers to early stage PBC disease as defined by the Rotterdam criteria.)

In addition, as it relates to the second of the two significant 747-301 inclusion criteria, a large majority of enrolled patients were also being administered concomitant UDCA. Hence, as a whole, the enrolled 747-301 trial population primarily consisted of early stage PBC patients who were being administered UDCA (see Section 3.2.3 below). Due to this circumstance, all of the previously presented biochemical marker cut points purported to be prognostic of transplant-

free survival (including what was ultimately chosen by Intercept) could not be appropriately applied (and hence even be adjudicated by DGIEP as a surrogate reasonably likely to predict clinical benefit) to assess treatment on this very specific PBC patient population that was enrolled in the applicant's registration trial.

Consequently, a biochemical marker cut point applicable only to the 747-301 enrolled patient population needed to be developed/studied during this NDA review cycle; this specifically was the basis for initiating the Major Amendment to the NDA review cycle as previously stated. This research need was met by a joint effort between the Global PBC Group and the DGIEP statistical review team. The Global PBC Study subject-level datasets were submitted to DGIEP with the goal of establishing a new criterion solely utilizing ALP reduction alone after 12 months of observation to better predict transplant-free survival within a subset of the Global PBC Study subjects that was comparable to the majority of enrolled patients in study 747-301 (i.e., early stage disease with concomitant UDCA usage). The determination as to whether ALP reduction alone (after 12 months of observation in early stage patients who are also using concomitant UDCA) is reasonably likely to predict clinical benefit was critically dependent on this joint effort. The corresponding overall statistical evaluation of ALP reduction in the aforementioned Global PBC Study patient subset is described within Dr. Min's final review document; please see her review, along with the clinical review as well, for full details. One of the best performing cut points, which resulted from Dr. Min's analysis along with subsequent follow-up conversations with the clinical review team in DGIEP, is utilized below (see Section 3.2.4) for conducting appropriate re-analyses of the 747-301 trial data.

3.2.2 Statistical Methodologies

Analysis Sets

The primary analysis set used for all efficacy analyses, along with the summarization of disposition along with demographics and baseline characteristics, was the 'Intent-to-Treat' (ITT) analysis set. This analysis set included all randomized patients who received at least one dose of blinded study drug. When utilizing this analysis set, patients were analyzed according to the treatment group that they were randomized to receive regardless of the actual treatment received. As can be seen in Section 3.2.3 below, all but one randomized patient received at least one dose of blinded study drug.

For sensitivity analysis purposes, all efficacy analyses were repeated utilizing a 'Completer' analysis set. This analysis set was comprised of all ITT patients who participated through the end of the double-blind period (i.e. through the Month 12 visit). When utilizing this analysis set, patients were analyzed according to the treatment group that they were randomized to receive regardless of the actual treatment received.

For additional sensitivity analysis purposes, all efficacy analyses were again repeated utilizing an 'Efficacy Evaluable' (EE) analysis set. This analysis set was comprised of all 'Completer' patients who did not have any major protocol deviations that would potentially affect the efficacy of the study drug. This analysis set definition was finalized in a blinded manner prior to database lock.

15

Multiplicity Adjustment

In order to control the overall study-wise type I error rate, a step-down/closed sequential testing procedure was pre-specified by the applicant to adjust for the multiple comparisons of the two OCA dose groups individually to placebo on the primary study endpoint alone. Starting with the 10 mg OCA comparison to placebo on the primary endpoint, the applicant stated that the step-down could only be carried to the OCA Titration comparison to placebo (on the primary endpoint), if and only if the 10 mg OCA comparison to placebo was found to be statistically significant (i.e., p-value less than 0.05). If the 10 mg OCA comparison to placebo was not statistically significant (i.e., p-value greater than or equal to 0.05), then the hypothesis test for the OCA Titration comparison to placebo on the primary endpoint would be deemed as exploratory.

As can be deduced, this pre-specified multiplicity adjustment procedure was narrow in scope in that it only pertained to the individual OCA dose comparisons with placebo on the primary endpoint alone. Hence even if both OCA dose comparisons were found to be statistically significant, then any other hypothesis test would still be deemed as exploratory in nature.

Primary Endpoint Analysis

The primary composite endpoint was assessed for patients within the OCA and placebo treatment groups. For descriptive purposes, the responder rates at Months 6 and 12 were calculated for all treatment groups separately along with corresponding 95% Wald Confidence Intervals (CI).

The applicant's analysis, based on DGIEP advice, utilized a Cochran-Mantel-Haenszel (CMH) test which adjusted for both randomization stratification variables (as previously described above in Section 3.2.1). In tandem with the CMH test, a Breslow-Day test was also conducted in order to test for the homogeneity of the treatment effect across the different randomization strata.

Descriptive Supportive Analyses

There were no formal secondary endpoints. As stated previously, other trial objectives such as assessing safety, histological, bile acid, and biomarker (i.e., not including ALP and TB) parameters were considered exploratory from a statistical perspective and hence are not presented in this review (please see the clinical review document for full details).

Several descriptive analyses were presented by the statistical reviewer to further support the primary endpoint analysis. These pertained to descriptively assessing the absolute change-from-baseline and percentage change-from-baseline in ALP and TB concentrations at Month 12 separately for each treatment group. The proportion of patients achieving a decrease in ALP of at least 10%, 15%, 20%, and 40% at Month 12 was presented separately for each treatment group. In addition, proportions for shift from baseline at Month 12 pertaining to relevant ALP and TB multiples of ULN categories were presented separately for each treatment group. Finally, separate figures presenting ALP and TB concentrations over time were presented to assess the durability of biochemical response while continuing long term treatment.

Handling of Dropouts/Missing Data

To assess the sensitivity of the results to missing/unavailable Month 12 data, a worst-case (i.e., designating "failure") imputation strategy was espoused by the applicant for the primary

16

endpoint analyses. An additional ultra-worst-case imputation strategy was espoused by the statistical reviewer for the same analyses; this new strategy imputed "failure" at Month 12 for OCA treated patients having missing/unavailable Month 12 data while imputing "success" at Month 12 for placebo treated patients having missing/unavailable data at Month 12. As is discussed in Section 3.2.4 below, the final results and conclusions were not influenced by the limited missing data encountered in the study.

Other Analysis Considerations
Baseline was defined by the applicant as the last measurement prior to the first administration of study drug, or, if multiple pre-treatment measurements were available, the arithmetic mean of the last (up to) three measurements preceding the first administration of study drug. Unscheduled measurements prior to first study drug administration were considered in the calculation of baseline value.

For sensitivity analysis purposes, all relevant analyses were re-conducted by the statistical reviewer utilizing the median (in lieu of the arithmetic mean) of the pre-first dose measurements. In addition, all relevant analyses were again re-conducted by the statistical reviewer utilizing a traditional baseline definition, i.e., choosing the last non-missing value prior to the first administration of study drug.

As stated previously, the enrolled 747-301 trial population primarily consisted of early stage PBC patients who also had screening/baseline ALP $\geq 1.67 \times$ULN and who were being administered UDCA (see Section 3.2.3 below). Due to this circumstance, the applicant's pre-specified primary composite endpoint could not be appropriately applied to assess treatment on this very specific PBC patient population that was enrolled in this registration trial. The goal to establish a new criterion solely utilizing ALP reduction alone after 12 months of observation to better predict transplant-free survival within a subset of the Global PBC Study subjects that was comparable to the majority of enrolled patients in study 747-301 was achieved by Dr. Min. Hence the previously described analyses, if applicable/relevant, were repeated by the statistical reviewer (analyzing trial patients who were exclusively early stage PBC patients, who also had screening/baseline ALP $\geq 1.67 \times$ULN while being administered UDCA) utilizing this newly determined, and relatively best performing, criterion (see Section 3.2.4 below for details). Other relevant ALP cut points explored by Dr. Min in her review were also applied to this subset of trial patients for sensitivity analysis purposes.

3.2.3 Patient Disposition and Demographic and Baseline Characteristics

Prior to displaying the information pertaining to patient disposition along with demographic and baseline characteristics, it should be noted that 217 patients were enrolled and randomized into the 747-301 study with 216 being administered at least one dose of study drug (i.e., part of the ITT analysis set as defined above). The lone patient who was randomized but not dosed was from the OCA Titration treatment group; 71 patients were enrolled and randomized to the OCA Titration treatment group with 70 of these patients being administered at least one dose of study medication (the patient who dropped out immediately after randomization withdrew consent). With one OCA Titration patient discontinuing prior to Month 6, a total of 37 out of 69 OCA Titration patients were eligible for up-titration at Month 6. Ultimately 33 of these 37 eligible

17

patients were titrated up to the 10 mg dose; hence 36 of the 69 OCA Titration patients at Month 6 remained on 5 mg OCA.

The disposition information for all ITT patients is displayed in Table 2 below.

Table 2
Disposition
(ITT)

	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)	Total (N = 216)
Intent-to-Treat (ITT)	73 (100%)	70 (100%)	73 (100%)	216 (100%)
Completer	64 (87.7%)	64 (91.4%)	70 (95.9%)	198 (91.7%)
Efficacy Evaluable (EE)	63 (86.3%)	63 (90.0%)	67 (91.8%)	193 (89.4%)
Discontinued Study Early	9 (12.3%)	6 (8.6%)	3 (4.1%)	18 (8.3%)
Death	0	1 (1.4%)	0	1 (0.5%)
Pruritus	7 (9.6%)	1 (1.4%)	0	8 (3.7%)
Other Adverse Events (AEs)	1 (1.4%)	3 (4.3%)	2 (2.7%)	6 (2.8%)
Withdrew Consent	1 (1.4%)	1 (1.4%)	1 (1.4%)	3 (1.4%)
Participated in LTSE	64 (87.7%)	63 (90.0%)	66 (90.4%)	193 (89.4%)

Source: Reviewer's Table generated from the 747-301 ADSL dataset.
Note: Denominators for percentages are N.

The relevant demographics and baseline characteristics for all ITT patients are presented in Table 3 below.

18

Table 3
Demographic and Baseline Characteristics
(ITT)

	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)	Total (N = 216)
Age at Screening (years)				
n	73	70	73	216
Mean (SD)	56.2 (11.00)	55.8 (10.52)	55.5 (10.03)	55.8 (10.48)
Median	56.0	54.5	55.0	55.0
Min, Max	30, 86	29, 83	35, 78	29, 86
Age Category – n (%)				
< 65 years old	56 (76.7%)	60 (85.7%)	60 (82.2%)	176 (81.5%)
≥ 65 years old	17 (23.3%)	10 (14.3%)	13 (17.8%)	40 (18.5%)
PBC Diagnosis Age (years)				
n	73	70	73	216
Mean (SD)	47.1 (10.60)	47.6 (11.65)	47.3 (9.34)	47.3 (10.51)
Median	47.0	48.0	48.0	47.5
Min, Max	24, 78	25, 82	31, 74	24, 82
PBC Diagnosis Age Category – n (%)				
< 45 years old	28 (38.4%)	29 (41.4%)	29 (39.7%)	86 (39.8%)
≥ 45 years old	45 (61.6%)	41 (58.6%)	44 (60.3%)	130 (60.2%)
Diagnosis Year Category – n (%)				
< 1990	4 (5.5%)	2 (2.9%)	0	6 (2.8%)
≥ 1990	69 (94.5%)	68 (97.1%)	73 (100%)	210 (97.2%)
Duration of PBC (years)				
n	73	70	73	216
Mean (SD)	9.2 (6.85)	8.3 (5.79)	8.3 (5.39)	8.6 (6.03)
Median	8.5	7.2	7.4	7.8
Min, Max	0.04, 32	0.3, 27	0.9, 22	0.04, 32
Duration of PBC Category – n (%)				
< 7.5 years	30 (41.1%)	36 (51.4%)	39 (53.4%)	105 (48.6%)
≥ 7.5 years	43 (58.9%)	34 (48.6%)	34 (46.6%)	111 (51.4%)
Gender – n (%)				
Female	63 (86.3%)	65 (92.9%)	68 (93.2%)	196 (90.7%)
Male	10 (13.7%)	5 (7.1%)	5 (6.9%)	20 (9.3%)
Race – n (%)				
Asian	1 (1.4%)	1 (1.4%)	1 (1.4%)	3 (1.4%)
Black or African American	1 (1.4%)	1 (1.4%)	1 (1.4%)	3 (1.4%)
Other	1 (1.4%)	1 (1.4%)	5 (6.9%)	7 (3.2%)
White	70 (95.9%)	67 (95.7%)	66 (90.4%)	203 (94.0%)
Geographical Region – n (%)				
Australia	1 (1.4%)	5 (7.1%)	3 (4.1%)	9 (4.2%)
Europe	51 (69.9%)	45 (64.3%)	49 (67.1%)	145 (67.1%)
North America	21 (28.8%)	20 (28.6%)	21 (28.8%)	62 (28.7%)

Source: Reviewer's Table generated from the 747-301 ADSL and ADLIVER datasets.
Note: Denominators for percentages are N.

19

Table 3 continued:
Demographic and Baseline Characteristics
(ITT)

	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)	Total (N = 216)
UDCA Usage – n (%)				
Yes	67 (91.8%)	65 (92.9%)	68 (93.2%)	200 (92.6%)
No	6 (8.2%)	5 (7.1%)	5 (6.9%)	16 (7.4%)
Total Daily UDCA Dose (mg)				
n	67	65	68	200
Mean (SD)	1110.5 (328.40)	1116.2 (289.41)	1061.8 (322.43)	1095.8 (313.55)
Median	1000.0	1000.0	1000.0	1000.0
Min, Max	300, 2000	600, 1800	500, 2700	300, 2700
Randomization Strata – n (%)				
1. ALP ≤ 3×ULN and AST ≤ 2×ULN and TB ≤ ULN; UDCA Usage	45 (61.6%)	43 (61.4%)	45 (61.6%)	133 (61.6%)
2. ALP ≤ 3×ULN and AST ≤ 2×ULN and TB ≤ ULN; No UDCA Usage	2 (2.7%)	2 (2.9%)	2 (2.7%)	6 (2.8%)
3. ALP > 3×ULN and/or AST > 2×ULN and/or TB > ULN; UDCA Usage	23 (31.5%)	22 (31.4%)	23 (31.5%)	68 (31.5%)
4. ALP > 3×ULN and/or AST > 2×ULN and/or TB > ULN; No UDCA Usage	3 (4.1%)	3 (4.3%)	3 (4.1%)	9 (4.2%)
ALP Concentration (U/L)				
n	73	70	73	216
Mean (SD)	316.3 (103.88)	325.9 (116.24)	327.5 (115.01)	323.2 (111.37)
Median	271.3	281.3	311.9	286.6
Min, Max	207, 620	187, 811	144, 746	144, 811
ALP Concentration (×ULN)				
n	73	70	73	216
Mean (SD)	2.658 (0.878)	2.747 (0.9851)	2.760 (0.9732)	2.721 (0.9431)
Median	2.293	2.378	2.607	2.423
Min, Max	1.68, 5.23	1.58, 6.86	1.22, 6.31	1.22, 6.86
TB Concentration (μmol/L)				
n	73	70	73	216
Mean (SD)	11.3 (6.69)	10.3 (5.51)	11.8 (7.38)	11.1 (6.59)
Median	9.2	9.1	9.2	9.1
Min, Max	2, 34	2, 36	2, 39	2, 39
TB Concentration (×ULN)				
n	73	70	73	216
Mean (SD)	0.558 (0.3162)	0.514 (0.2490)	0.598 (0.3733)	0.557 (0.3181)
Median	0.473	0.456	0.478	0.469
Min, Max	0.08, 1.78	0.11, 1.43	0.12, 2.03	0.08, 2.03

Source: Reviewer's Table generated from the 747-301 ADSL and ADLIVER datasets.
Note: Denominators for percentages are N.

20

Table 3 continued:
Demographic and Baseline Characteristics
(ITT)

	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)	Total (N = 216)
PBC Disease Stage using Rotterdam Criteria				
Early	65 (89.0%)	64 (91.4%)	66 (90.4%)	195 (90.3%)
Moderately Advanced	8 (11.0%)	6 (8.6%)	7 (9.6%)	21 (9.7%)
Albumin ≤ LLN	1 (1.4%)	2 (2.9%)	0	3 (1.4%)
TB > 1.0×ULN	7 (9.6%)	4 (5.7%)	7 (9.6%)	18 (8.3%)
Advanced	0	0	0	0
ALP Baseline Categories – n (%)				
1. 1.0×ULN < ALP < 1.67×ULN	0	1 (1.4%)	1 (1.4%)	2 (0.9%)
2. 1.67×ULN ≤ ALP < 2.0×ULN	20 (27.4%)	13 (18.6%)	16 (21.9%)	49 (22.7%)
3. 2.0×ULN ≤ ALP < 3.0×ULN	33 (45.2%)	37 (52.9%)	33 (45.2%)	103 (47.7%)
4. 3.0×ULN ≤ ALP < 4.0×ULN	12 (16.4%)	10 (14.3%)	15 (20.5%)	37 (17.1%)
5. 4.0×ULN ≤ ALP < 5.0×ULN	6 (8.2%)	8 (11.4%)	5 (6.8%)	19 (8.8%)
6. ALP ≥ 5.0×ULN	2 (2.7%)	1 (1.4%)	3 (4.1%)	6 (2.8%)
TB Baseline Categories – n (%)				
1. TB ≤ 1.0×ULN	66 (90.4%)	66 (94.3%)	66 (90.4%)	198 (91.7%)
2. 1.0×ULN < TB < 2.0×ULN	7 (9.6%)	4 (5.7%)	6 (8.2%)	17 (7.8%)
3. TB ≥ 2.0×ULN	0	0	1 (1.4%)	1 (0.5%)
Relevant Combination Baseline Categories – n (%)				
1. ALP ≥ 1.67×ULN and Early Stage PBC Disease; UDCA Usage	**60 (82.2%)**	**60 (85.7%)**	**61 (83.6%)**	**181 (83.8%)**
2. 1.67×ULN < ALP < 2.0×ULN and Early Stage PBC Disease; UDCA Usage	18 (24.7%)	13 (18.6%)	15 (20.5%)	46 (21.3%)
3. ALP ≥ 2.0×ULN and Early Stage PBC Disease; UDCA Usage	42 (57.5%)	47 (67.1%)	46 (63.0%)	135 (62.5%)

Source: Reviewer's Table generated from the 747-301 ADSL and ADLIVER datasets.
Note: Denominators for percentages are N.

Note that patients 105006 (Placebo) and 165002 (OCA Titration) were the only two patients in the study with baseline ALP < 1.67×ULN. And one of these patients (i.e., 165002) also had a normal baseline total bilirubin concentration as well. It can be seen from the presented demographic and baseline characteristics that there was balance, for all presented variables, between the randomized treatment groups.

3.2.4 Results and Conclusions

All results presented in this section were generated by the statistical reviewer.

Table 4
Proportion of Patients who Achieved Response
(ITT)

Statistics	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)
Response at Month 6 – n (%) [1] [2]	37 (50.7%)	24 (34.3%)	5 (6.9%)
Corresponding 95% Wald CI	39.2%, 62.2%	23.2%, 45.4%	1.1%, 12.6%
Response at Month 12 – n (%) [1] [2]	34 (46.6%)	32 (45.7%)	7 (9.6%)
Corresponding 95% Wald CI	36.5%, 59.4%	34.0%, 57.4%	2.8%, 16.3%
CMH Test p-value [3]	<0.0001	<0.0001	
Corresponding Breslow-Day Test p-value	0.9072	0.5045	
(1) ALP < 1.67×ULN at Month 12 – n (%) [2]	40 (54.8%)	33 (47.1%)	12 (16.4%)
(2) TB ≤ 1.0×ULN at Month 12 – n (%) [2]	60 (82.2%)	62 (88.6%)	57 (78.1%)
(3) Decrease in ALP ≥ 15% at Month 12 – n (%) [2]	57 (78.1%)	54 (77.1%)	21 (28.8%)

Source: Reviewer's Table generated from ADLIVER dataset.
Note: Denominators for percentages are N.
[1]: A patient was designated as a responder if all three of the following conditions were met: (1) 12-Month value of ALP < 1.67×ULN; (2) 12-Month value of TB ≤ ULN; (3) ALP reduction from baseline at Month 12 ≥ 15%.
[2]: Patients with missing data at these timepoints were designated as non-responders.
[3]: Month 12 Pair-wise comparison made between given OCA treatment group and Placebo adjusted for both randomization stratification variables.

It can be observed from Table 4 above that both OCA treatment groups showed a superior difference in the proportion/percentage of patients achieving response at Month 12 when individually compared to placebo using the CMH test. The corresponding Breslow-Day test result shows that the treatment effects were homogeneous across the different randomization strata. This analysis was repeated utilizing the Completer and EE analysis sets and the conclusions were consistent. The ultra-worse-case imputation strategy, implemented by the statistical reviewer as described above in Section 3.2.2, did not impact the study conclusions. It is important to note that no single site influenced or drove the overall study results. In regards to ALP or TB values at Month 12, there were no patients who were designated as outliers (i.e., by having studentized residual values greater than three), and there was no impact on study conclusions between corrected laboratory values (as presented) and original (i.e., uncorrected) laboratory values. All of the previously presented analyses were re-conducted utilizing a baseline value that was the median of all pre-first dose measurements, and, separately, a traditional baseline definition (both approaches as described earlier in Section 3.2.2 above); there was no impact on study conclusions with either approach. Considering the applicant's pre-specified step-down/closed sequential testing procedure as previously described in Section 3.2.2, formal hypothesis testing is stopped at this point. Any subsequent inferential statistic reported below should be considered exploratory.

Table 5
ALP Summary at Month 12
(ITT)

Time Point/Statistics	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)
Baseline ALP Concentration (U/L)			
n	73	70	73
Mean (SD)	316.3 (103.88)	325.9 (116.24)	327.5 (115.01)
Median	271.3	281.3	311.9
Min, Max	207, 620	187, 811	144, 746
Month 12 ALP Concentration (U/L)			
n	63	64	70
Mean (SD)	191.2 (61.38)	219.5 (99.76)	321.3 (142.88)
Median	181.7	196.6	270.5
Min, Max	95, 444	116, 690	149, 733
Absolute Change from Baseline to Month 12 (U/L)			
n	63	64	70
Mean (SD)	-117.1 (72.84)	-103.5 (87.03)	-7.7 (87.96)
Median	-99.0	-85.5	-15.8
Min, Max	-346, 0.3	-402, 127	-208, 308
Percentage Change from Baseline to Month 12 (%)			
n	63	64	70
Mean (SD)	-36.4 (14.88)	-30.5 (18.97)	-2.5 (23.82)
Median	-38.3	-31.5	-4.7
Min, Max	-72, 0.1	-74, 23	-45, 80
Decrease in ALP ≥ 10% at Month 12 – n (%) [1]	61 (83.6%)	55 (78.6%)	29 (39.7%)
Decrease in ALP ≥ 20% at Month 12 – n (%) [1]	54 (74.0%)	49 (70.0%)	17 (23.3%)
Decrease in ALP ≥ 40% at Month 12 – n (%) [1]	25 (34.3%)	21 (30.0%)	1 (1.4%)
Baseline ALP Concentration (×ULN)			
n	73	70	73
Mean (SD)	2.658 (0.878)	2.747 (0.9851)	2.760 (0.9732)
Median	2.293	2.378	2.607
Min, Max	1.68, 5.23	1.58, 6.86	1.22, 6.31
Month 12 ALP Concentration (×ULN)			
n	63	64	70
Mean (SD)	1.606 (0.5161)	1.851 (0.8449)	2.705 (1.1987)
Median	1.527	1.661	2.286
Min, Max	0.80, 3.75	0.98, 5.84	1.26, 6.19

Source: Reviewer's Table generated from ADLIVER dataset.
Note: Denominators for percentages are N.
[1]: Patients with missing data at these timepoints were designated as non-responders.

23

Table 6
ALP Shift from Baseline Summary at Month 12
(ITT)

Time Point/Statistics	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)
Baseline ALP ≤ 1.0×ULN	0	0	0
1.0×ULN < Baseline ALP < 1.67×ULN	0	1 (1.4%)	1 (1.4%)
Baseline ALP ≥ 1.67×ULN	73 (100%)	69 (98.6%)	72 (98.6%)
1.0×ULN < Baseline ALP < 1.67×ULN	n=0	n=1	n=1
Month 12 ALP ≤ 1.0×ULN [1]	0	0	0
1.0×ULN < Month 12 ALP < 1.67×ULN [1]	0	1 (100%)	1 (100%)
Month 12 ALP ≥ 1.67×ULN [1]	0	0	0
Month 12 ALP missing [1]	0	0	0
Baseline ALP ≥ 1.67×ULN	n=73	n=69	n=72
Month 12 ALP ≤ 1.0×ULN [2]	5 (6.9%)	1 (1.5%)	0
1.0×ULN < Month 12 ALP < 1.67×ULN [2]	35 (48.0%)	31 (44.9%)	11 (15.3%)
Month 12 ALP ≥ 1.67×ULN [2]	23 (31.5%)	31 (44.9%)	58 (80.6%)
Month 12 ALP missing [2]	10 (13.7%)	6 (8.7%)	3 (4.2%)

Source: Statistical Reviewer's Table.
Note: Denominators for percentages are N.
[1]: The denominator for this calculation is the number of patients with 1.0×ULN < Baseline ALP < 1.67×ULN.
[2]: The denominator for this calculation is the number of patients with Baseline ALP ≥ 1.67×ULN.

It can be observed from Tables 5 and 6 above that both OCA treatment arms reduced ALP relative to placebo. It should be noted that the continuous descriptive statistics pertaining to the baseline, Month 12, absolute change from baseline at Month 12 and percentage change from baseline at Month 12 values utilized only the available data at those time points (i.e., no missing data were imputed). The categorical descriptive statistics (i.e., frequencies and corresponding proportions at Month 12) utilized the worse-case imputation strategy. The applicant's baseline definition was used for all presented calculations.

24

Figure 2
Patient Profiles for ALP Concentration Change during 12 Month Double Blind Period by Baseline TB Category
(ITT)
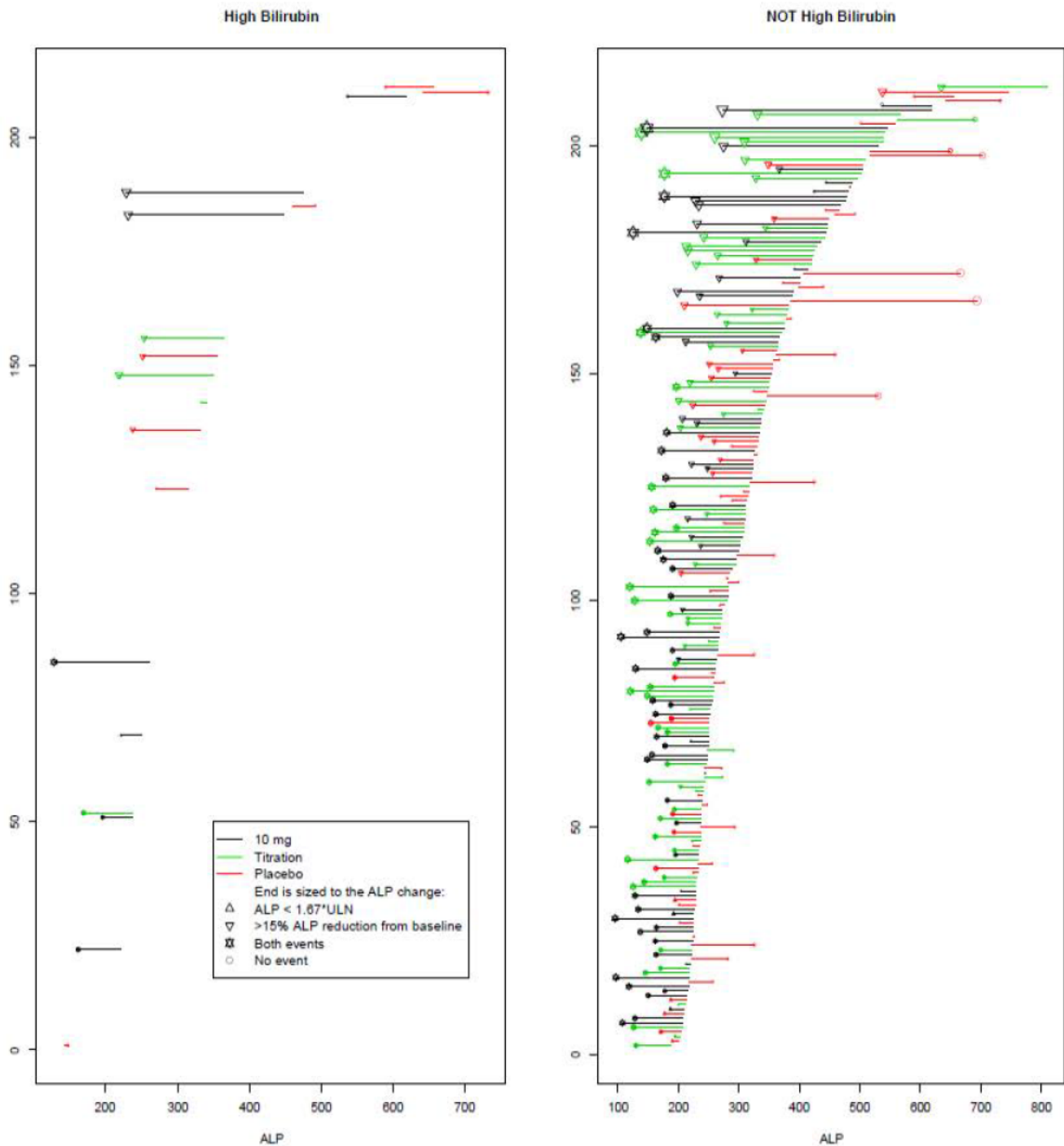
Source: Figure by Dr. Andrejus Parfionovas generated from ADLIVER dataset.

Figure 2 above presents the ALP concentration (in U/L) for each of the 216 ITT patients as it changes from baseline throughout the 12 month double-blind treatment period; the length of a

25

line represents the magnitude of change in ALP. This graphical patient profiles presentation is made by baseline TB category, in that the left panel presents the 18 patients with baseline TB > 1.0×ULN while the right panel presents the 198 patients with baseline TB ≤ 1.0×ULN. This figure dovetails with Table 5 and 6 above to represent each OCA treatment group's effectiveness in reducing ALP concentration levels relative to placebo.

After completing the 12-month double-blind treatment period, 193 out of the 216 ITT patients (i.e., 64 10 mg OCA, 63 OCA Titration, and 66 Placebo patients) continued on open-label OCA treatment during the LTSE period. Figure 3 below presents ALP concentration over time for all ITT patients, organized by originally randomized treatment group, through the currently ongoing open-label LTSE period up to the latest data cut made on June 29, 2015.

Figure 3
ALP Concentration (U/L) from Randomization through Latest LTSE Data Cut – 1
(ITT)
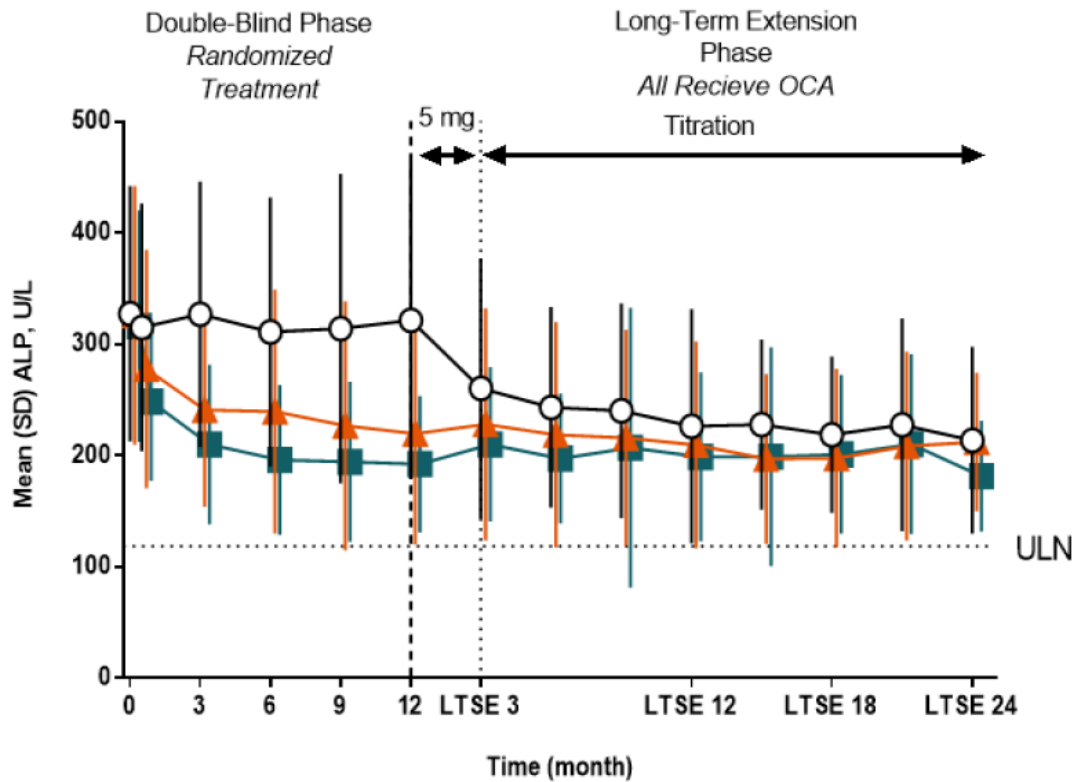
Month	0	0.5	3	6	9	12	LTSE 3	LTSE 6	LTSE 9	LTSE 12	LTSE 15	LTSE 18	LTSE 21	LTSE 24
Placebo n	73	72	69	71	69	70	64	60	59	59	55	55	28	12
Titration OCA n	70	68	69	69	66	64	63	62	62	60	57	56	29	10
OCA 10 mg n	73	71	66	64	64	62	64	59	61	59	58	59	30	12

Source: Figure 3 of page 13 of the 120 Day Safety Update submitted on October 30, 2015 (eCTD sequence 0034).

It can again be seen that ALP concentration levels are reduced by both OCA treatment groups during the first 12 months, most notably during the first three months; these reduced levels

26

remained stable during the LTSE period suggesting durability of response. It can also be seen that ALP levels for placebo patients were flat during the first 12 months; however, these levels started decreasing immediately, and ultimately remained stable, during the LTSE period once these patients started OCA administration.

Table 7
TB Summary at Month 12
(ITT)

Time Point/Statistics	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)
Baseline TB Concentration (µmol/L)			
n	73	70	73
Mean (SD)	11.3 (6.69)	10.3 (5.51)	11.8 (7.38)
Median	9.2	9.1	9.2
Min, Max	2, 34	2, 36	2, 39
Month 12 TB Concentration (µmol/L)			
n	63	64	70
Mean (SD)	9.6 (4.68)	9.9 (4.82)	13.2 (8.69)
Median	7.9	8.2	9.8
Min, Max	2, 25	4, 28	4, 45
Absolute Change from Baseline to Month 12 (µmol/L)			
n	63	64	70
Mean (SD)	-1.2 (4.36)	-0.62 (3.33)	1.4 (4.13)
Median	-0.46	-0.34	1.3
Min, Max	-18, 7	-9, 7	-7, 20
Percentage Change from Baseline to Month 12 (%)			
n	63	64	70
Mean (SD)	-1.1 (36.19)	1.3 (34.71)	17.0 (41.54)
Median	-5.1	-5.0	12.4
Min, Max	-51, 194	-51, 125	-43, 211
Baseline TB Concentration (×ULN)			
n	73	70	73
Mean (SD)	0.558 (0.3162)	0.514 (0.2490)	0.598 (0.3733)
Median	0.473	0.456	0.478
Min, Max	0.08, 1.78	0.11, 1.43	0.12, 2.03
Month 12 TB Concentration (×ULN)			
n	63	64	70
Mean (SD)	0.479 (0.2332)	0.496 (0.2221)	0.660 (0.4097)
Median	0.407	0.416	0.496
Min, Max	0.12, 1.28	0.22, 1.12	0.23, 1.96

Source: Reviewer's Table generated from ADLIVER dataset.
Note: Denominators for percentages are N.

27

Table 8
TB Shift from Baseline Summary at Month 12
(ITT)

Time Point/Statistics	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)
Baseline TB ≤ 1.0×ULN	66 (90.4%)	66 (94.3%)	66 (90.4%)
Baseline TB > 1.0×ULN	7 (9.6%)	4 (5.7%)	7 (9.6%)
Baseline TB ≤ 1.0×ULN	n=66	n=66	n=66
Month 12 TB ≤ 1.0×ULN [1]	55 (83.3%)	60 (90.1%)	56 (84.4%)
Month 12 TB > 1.0×ULN [1]	3 (4.5%)	0	7 (10.6%)
Month 12 TB missing [1]	8 (12.1%)	6 (9.1%)	3 (4.5%)
Baseline TB > 1.0×ULN	n=7	n=4	n=7
Month 12 TB ≤ 1.0×ULN [2]	5 (71.4%)	2 (50.0%)	1 (14.2%)
Month 12 TB > 1.0×ULN [2]	0	2 (50.0%)	6 (85.7%)
Month 12 TB missing [2]	2 (28.6%)	0	0

Source: Statistical Reviewer's Table.
Note: Denominators for percentages are N.
[1]: The denominator for this calculation is the number of patients with TB ≤ 1.0×ULN at Baseline.
[2]: The denominator for this calculation is the number of patients with TB > 1.0×ULN at Baseline.

As observed from Tables 7 and 8 above, reductions from baseline in TB were greater in both OCA treatment groups than in the placebo group. However, note that very few patients had elevations in TB above ULN at baseline. It should be additionally noted that of these 18 patients with baseline elevations in TB, 2 of 7 (28.6%) in the OCA 10 mg arm, 1 of 4 (25.0%) in the OCA Titration arm and 0 of 7 in the placebo arm were designated as overall responders at Month 12 as pertaining to the pre-specified primary composite endpoint. It should also be noted that the continuous descriptive statistics pertaining to the baseline, Month 12, absolute change from baseline at Month 12 and percentage change from baseline at Month 12 values utilized only the available data at those time points (i.e., no missing data were imputed). The applicant's baseline definition was used for all presented calculations.

Figure 4 below presents TB concentration over time for all ITT patients, organized by originally randomized treatment group, through the currently ongoing open-label LTSE period up to the latest data cut made on June 29, 2015.

Figure 4
TB Concentration (µmol/L) from Randomization through Latest LTSE Data Cut – 1
(ITT)

Month	0	0.5	3	6	9	12	LTSE 3	LTSE 6	LTSE 9	LTSE 12	LTSE 15	LTSE 18	LTSE 21	LTSE 24
Placebo ± UDCA, n	73	72	68	71	69	70	64	60	59	59	55	55	28	12
Titration OCA ± UDCA, n	70	68	68	69	66	64	63	62	62	60	57	56	29	10
10 mg OCA ± UDCA, n	73	71	66	64	64	62	64	59	61	59	58	59	30	12

Source: Figure 7 of page 17 of the 120 Day Safety Update submitted on October 30, 2015 (eCTD sequence 0034).

It can be seen that reductions from baseline in TB were only marginally greater in both OCA treatment groups relative to the placebo group. It can also be seen that TB levels for placebo patients slightly increased during the first 12 months; however, these levels started decreasing immediately, and ultimately remained stable, during the LTSE period once these patients started OCA administration.

As stated previously, the enrolled 747-301 trial population primarily consisted of early stage PBC patients who also had screening/baseline ALP ≥ 1.67×ULN and who were being administered UDCA (see **bold** Relevant Combination Baseline Category 1 in Table 3 above). The goal to establish a new criterion solely utilizing ALP reduction alone after 12 months of observation to better predict transplant-free survival within a subset of the Global PBC Study subjects that was comparable to the majority of enrolled patients in study 747-301 (i.e., the 181 total patients as described above) was achieved by Dr. Min. This comparable Global PBC Study subset (i.e., early stage PBC subjects with screening/baseline ALP ≥ 1.67×ULN who were being

29

administered UDCA) consisted of 909 out of the original 4,845 subjects. The relevant demographics and baseline characteristics comparing these non-concurrent cohorts (i.e., the 181 patients from study 747-301 and the 909 subjects from the Global PBC Study) are presented in Table 9 below.

Table 9
Demographic and Baseline Characteristics of the Comparable Cohorts from Study 747-301 and the Global PBC Study

	Study 747-301 (N = 181)	Global PBC Study (N = 909)
Age at Screening (years)		
n	181	909
Mean (SD)	55.5 (9.82)	54.4 (11.16)
Median	54.0	54.0
Min, Max	29, 81	24, 86
Age Category – n (%)		
< 65 years old	151 (83.4%)	730 (80.3%)
≥ 65 years old	30 (16.6%)	179 (19.7%)
PBC Diagnosis Age (years)		
n	181	909
Mean (SD)	47.1 (10.03)	52.9 (11.24)
Median	47.0	53.0
Min, Max	25, 78	23, 86
PBC Diagnosis Age Category – n (%)		
< 45 years old	72 (39.8%)	209 (23.0%)
≥ 45 years old	109 (60.2%)	700 (77.0%)
Diagnosis Year Category – n (%)		
< 1990	2 (1.1%)	244 (26.8%)
≥ 1990	179 (98.9%)	665 (73.2%)
Duration of PBC (years)		
n	181	909
Mean (SD)	8.5 (5.63)	2.2 (3.79)
Median	7.8	0.27
Min, Max	0.4, 32	0, 36
Duration of PBC Category – n (%)		
< 7.5 years	87 (48.1%)	821 (90.3%)
≥ 7.5 years	94 (51.9%)	88 (9.7%)
Gender – n (%)		
Female	165 (91.2%)	842 (92.6%)
Male	16 (8.8%)	67 (7.4%)
Race – n (%)		
Asian	2 (1.1%)	Race
Black or African American	2 (1.1%)	Not
Other	6 (3.3%)	Available
White	171 (94.5%)	

Source: Reviewer's Table generated from the 747-301 ADSL and 747-301 ADLIVER datasets along with the GPBC_FDA and GPBClab_FDA datasets.
Note: Denominators for percentages are N. '*' signifies available Total Daily UDCA Dose data for 687 subjects. There was unavailable Total Daily UDCA Dose data for 202 subjects.

30

Table 9 continued:
**Demographic and Baseline Characteristics of the Comparable Cohorts from Study 747-301
and the Global PBC Study**

	Study 747-301 (N = 181)	Global PBC Study (N = 909)
Geographical Region – n (%)		
Australia	9 (5.0%)	0
Europe	118 (65.2%)	639 (70.3%)
North America	54 (29.8%)	270 (29.7%)
Total Daily UDCA Dose (mg)		
n	181	687*
Mean (SD)	1091.2 (312.66)	809.5 (233.66)
Median	1000.0	750.0
Min, Max	300, 2700	250, 1500
ALP Concentration (U/L)		
n	181	909
Mean (SD)	311.3 (95.54)	478.7 (390.77)
Median	281.5	388.0
Min, Max	200, 746	2, 2545
ALP Concentration (×ULN)		
n	181	909
Mean (SD)	2.621 (0.8101)	3.365 (1.770)
Median	2.380	2.722
Min, Max	1.68, 6.31	1.67, 15.30
TB Concentration (μmol/L)		
n	181	909
Mean (SD)	9.6 (4.37)	7.0 (5.65)
Median	8.3	8.0
Min, Max	2, 25	0.2, 22
TB Concentration (×ULN)		
n	181	909
Mean (SD)	0.480 (0.2077)	0.579 (0.2043)
Median	0.425	0.571
Min, Max	0.08, 0.99	0.12, 1.00

Source: Reviewer's Table generated from the 747-301 ADSL and 747-301 ADLIVER datasets along with the GPBC_FDA and GPBClab_FDA datasets.
Note: Denominators for percentages are N. '*' signifies available Total Daily UDCA Dose data for 687 subjects. There was unavailable Total Daily UDCA Dose data for 202 subjects.

It can be seen from Table 9 above that there were areas of imbalance; however, given the non-concurrent nature of these cohorts, the data were reasonably balanced. Notably there is a difference in disease duration between the two groups with the duration of disease from the Global PBC Study group being shorter. This may be secondary to the way the data was gathered and reported in the Global PBC Study, or may represent a real difference.

As presented within Dr. Min's review, many different cut point criteria that utilized ALP reduction alone after 12 months of observation for reasonably predicting transplant-free survival were explored and assessed within the 909 patient subset of the Global PBC Study. All of the

31

explored/assessed ALP cut points at 12 months were applied to the comparable 181 ITT patients from study 747-301 by treatment group for re-analysis purposes. The responder analysis results from the most relevant cut points explored are presented in Table 10 below.

Table 10
Proportion of Patients who Achieved Response at Month 12 by Relevant Explored ALP Cut Point Criteria
(Comparable ITT)

Explored Cut Points	10 mg OCA (N = 60)	OCA Titration (N = 60)	Placebo (N = 61)
ALP ≤ 1.0×ULN at Month 12 – n (%)	5 (8.3%)	1 (1.7%)	0
ALP < 1.67×ULN at Month 12 – n (%)	37 (61.7%)	29 (48.3%)	11 (18.0%)
ALP < 2.0×ULN at Month 12 – n (%)	47 (78.3%)	41 (68.3%)	20 (32.8%)
Decrease in ALP ≥ 40% at Month 12 – n (%)	19 (31.7%)	18 (30.0%)	1 (1.6%)
Decrease in ALP ≥ 15% at Month 12 – n (%)	48 (80.0%)	46 (76.7%)	19 (31.2%)
ALP < 1.67×ULN and Decrease ≥ 40% at Month 12 – n (%)	17 (28.3%)	12 (20.0%)	0
ALP < 1.67×ULN and Decrease ≥ 15% at Month 12 – n (%)	35 (58.3%)	28 (46.7%)	7 (11.5%)
ALP < 2.0×ULN and Decrease ≥ 40% at Month 12 – n (%)	18 (30.0%)	15 (25.0%)	1 (1.6%)
ALP < 2.0×ULN and Decrease ≥ 15% at Month 12 – n (%)	43 (71.7%)	36 (60.0%)	10 (16.4%)
Stratified Cut Point at Month 12 – n (%)	**26 (43.3%)**	**23 (38.3%)**	**3 (4.9%)**

Source: Reviewer's Table generated from ADLIVER dataset.
Note: Denominators for percentages are N.

It can be seen that applying all of these explored ALP cut points at 12 months resulted in consistent relative differences in response rates between the treatment groups. It should be noted that responder analysis results from ALP cut points assessed by Dr. Min that were not presented within Table 10 above were also consistent (i.e., similar relative differences in response rates between the treatment groups).

The stratified ALP cut point at Month 12 was defined as follows:
If baseline ALP was ≥ 2.0×ULN, then a patient would be designated as a responder if both of the following conditions were met:
- *12-Month value of ALP < 2.0×ULN*
- *ALP reduction from baseline at Month 12 ≥ 40%;*

Else if baseline ALP was ≥ 1.67×ULN but < 2.0×ULN, then a patient would be designated as a responder if both of the following conditions were met:
- *12-Month value of ALP < 1.67×ULN*
- *ALP reduction from baseline at Month 12 ≥ 15%.*

This stratified ALP cut point at Month 12 was the relatively better performing cut point per Dr. Min's analyses; please see her review for full details. An alteration of Table 4 above was reproduced (now as Table 11 below) by applying this stratified ALP cut point to the 181 comparable ITT patients for re-analysis purposes.

Table 11
Proportion of Patients who Achieved Response using Stratified Cut Point
(Comparable ITT)

Statistics	10 mg OCA (N = 60)	OCA Titration (N = 60)	Placebo (N = 61)
Response at Month 6 – n (%) [1]	**25 (41.7%)**	**21 (35.0%)**	**1 (1.6%)**
Corresponding 95% Wald CI	29.2%, 54.1%	22.9%, 47.1%	0.0%, 4.8%
Baseline ALP ≥ 2.0×ULN – n (%)	42 (70.0%)	47 (78.3%)	46 (75.4%)
ALP < 2.0×ULN at Month 6 – n (%) [2]	30 (71.4%)	24 (51.1%)	8 (17.4%)
Decrease in ALP ≥ 40% at Month 6 – n (%) [2]	10 (23.8%)	13 (27.7%)	0
ALP < 2.0×ULN and Decrease ≥ 40% at Month 6 – n (%) [2]	**9 (21.4%)**	**11 (23.4%)**	**0**
Baseline ALP ≥ 1.67×ULN but < 2.0×ULN – n (%)	18 (30.0%)	13 (21.7%)	15 (24.6%)
ALP < 1.67×ULN at Month 6 – n (%) [3]	17 (94.4%)	10 (76.9%)	3 (20.0%)
Decrease in ALP ≥ 15% at Month 6 – n (%) [3]	16 (88.9%)	11 (84.6%)	1 (6.7%)
ALP < 1.67×ULN and Decrease ≥ 15% at Month 6 – n (%) [3]	**16 (88.9%)**	**10 (76.9%)**	**1 (6.7%)**
Response at Month 12 – n (%) [1]	**26 (43.3%)**	**23 (38.3%)**	**3 (4.9%)**
Corresponding 95% Wald CI	30.8%, 55.9%	26.0%, 50.6%	0.0%, 10.3%
Baseline ALP ≥ 2.0×ULN – n (%)	42 (70.0%)	47 (78.3%)	46 (75.4%)
ALP < 2.0×ULN at Month 12 – n (%) [2]	29 (69.1%)	28 (59.6%)	9 (19.6%)
Decrease in ALP ≥ 40% at Month 12 – n (%) [2]	13 (31.0%)	16 (34.0%)	1 (2.2%)
ALP < 2.0×ULN and Decrease ≥ 40% at Month 12 – n (%) [2]	**12 (28.6%)**	**13 (27.7%)**	**1 (2.2%)**
Baseline ALP ≥ 1.67×ULN but < 2.0×ULN – n (%)	18 (30.0%)	13 (21.7%)	15 (24.6%)
ALP < 1.67×ULN at Month 12 – n (%) [3]	16 (88.9%)	11 (84.6%)	6 (40.0%)
Decrease in ALP ≥ 15% at Month 12 – n (%) [3]	14 (77.8%)	10 (76.9%)	2 (13.3%)
ALP < 1.67×ULN and Decrease ≥ 15% at Month 12 – n (%) [3]	**14 (77.8%)**	**10 (76.9%)**	**2 (13.3%)**

Source: Reviewer's Table generated from ADLIVER dataset.
Note: Denominators for percentages are N.
[1]: Response is defined by the Stratified ALP Cut Point.
[2]: The denominator for this calculation is the number of patients with Baseline ALP ≥ 2.0×ULN.
[3]: The denominator for this calculation is the number of patients with Baseline ALP ≥ 1.67×ULN but < 2.0×ULN.

It can be observed from Table 11 that both OCA treatment groups showed a difference in the proportion/percentage of patients achieving response at Month 12 when individually compared to placebo. This analysis was repeated utilizing the Completer and EE analysis sets and the conclusions were consistent. The ultra-worse-case imputation strategy, implemented by the statistical reviewer as described above in Section 3.2.2, did not impact the results. All of the previously presented analyses were re-conducted utilizing a baseline value that was the median of all pre-first dose measurements, and, separately, a traditional baseline definition (both

33

approaches as described earlier in Section 3.2.2 above); there was no impact on the results with either approach.

After completing the 12-month double-blind treatment period, 163 out of these 181 comparable ITT patients (i.e., 54 10 mg OCA, 53 OCA Titration, and 56 Placebo patients) continued on open-label OCA treatment during the LTSE period. Figure 5 below presents ALP concentration over time for the 181 comparable ITT patients, organized by originally randomized treatment group, through the currently ongoing open-label LTSE period up to the latest data cut made on June 29, 2015.

Figure 5
ALP Concentration (U/L) from Randomization through Latest LTSE Data Cut – 2
(Comparable ITT)
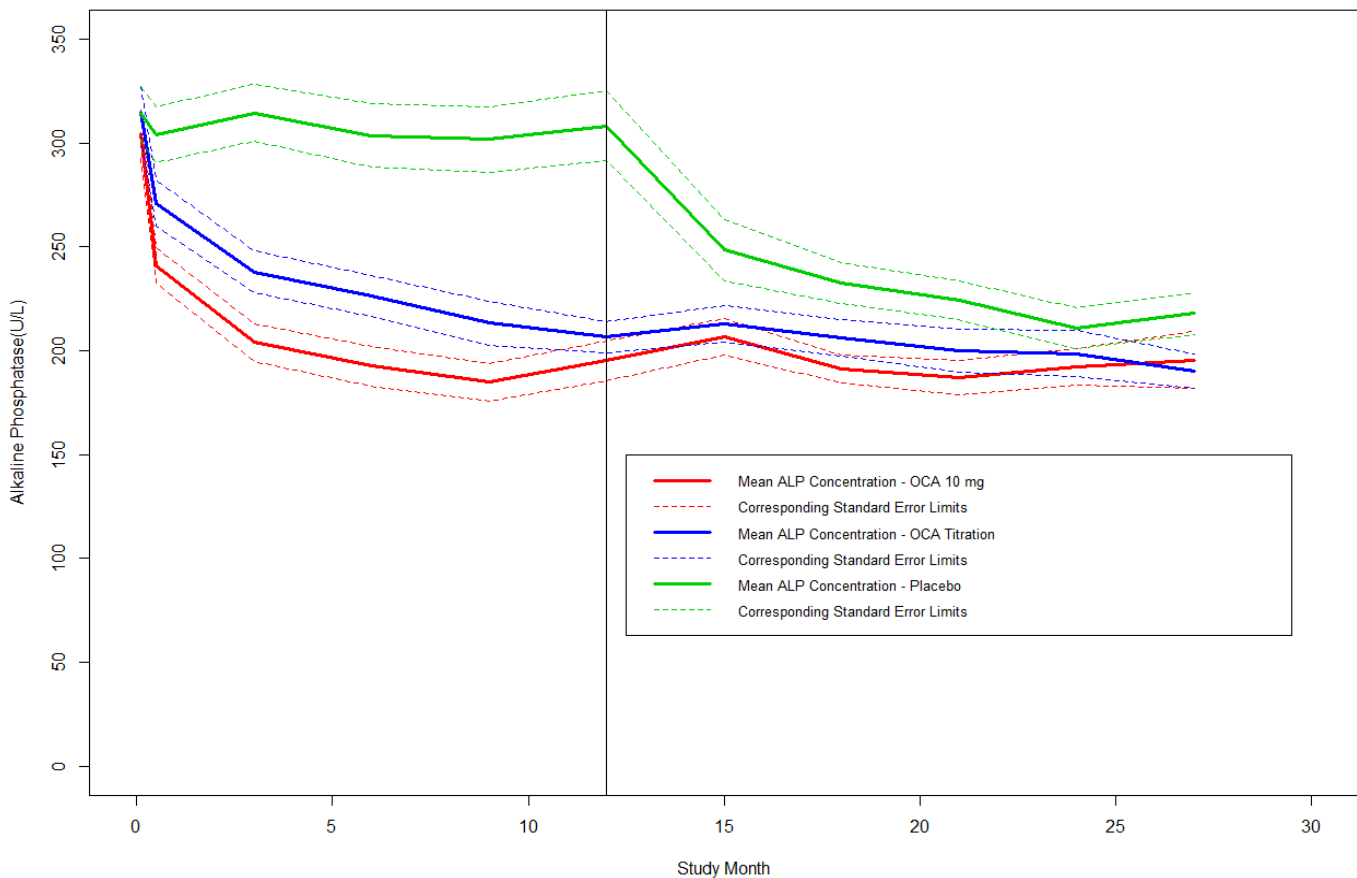
Source: Reviewer's Figure generated from integrated ADLB and ADLBOLDB datasets.

Similar to the whole ITT group results as previously presented in Figure 3, it can be seen that ALP concentration levels are reduced by both OCA treatment groups during the first 12 months, most notably during the first three months; these reduced levels remained stable during the LTSE period suggesting durability of response. It can also be seen that ALP levels for placebo patients were flat during the first 12 months; however, these levels started decreasing immediately, and ultimately remained stable, during the LTSE period once these patients started OCA administration.

34

Figure 6 below presents TB concentration over time for the 181 comparable ITT patients, organized by originally randomized treatment group, through the currently ongoing open-label LTSE period up to the latest data cut made on June 29, 2015.

Figure 6
TB Concentration (μmol/L) from Randomization through Latest LTSE Data Cut – 2 (ITT)
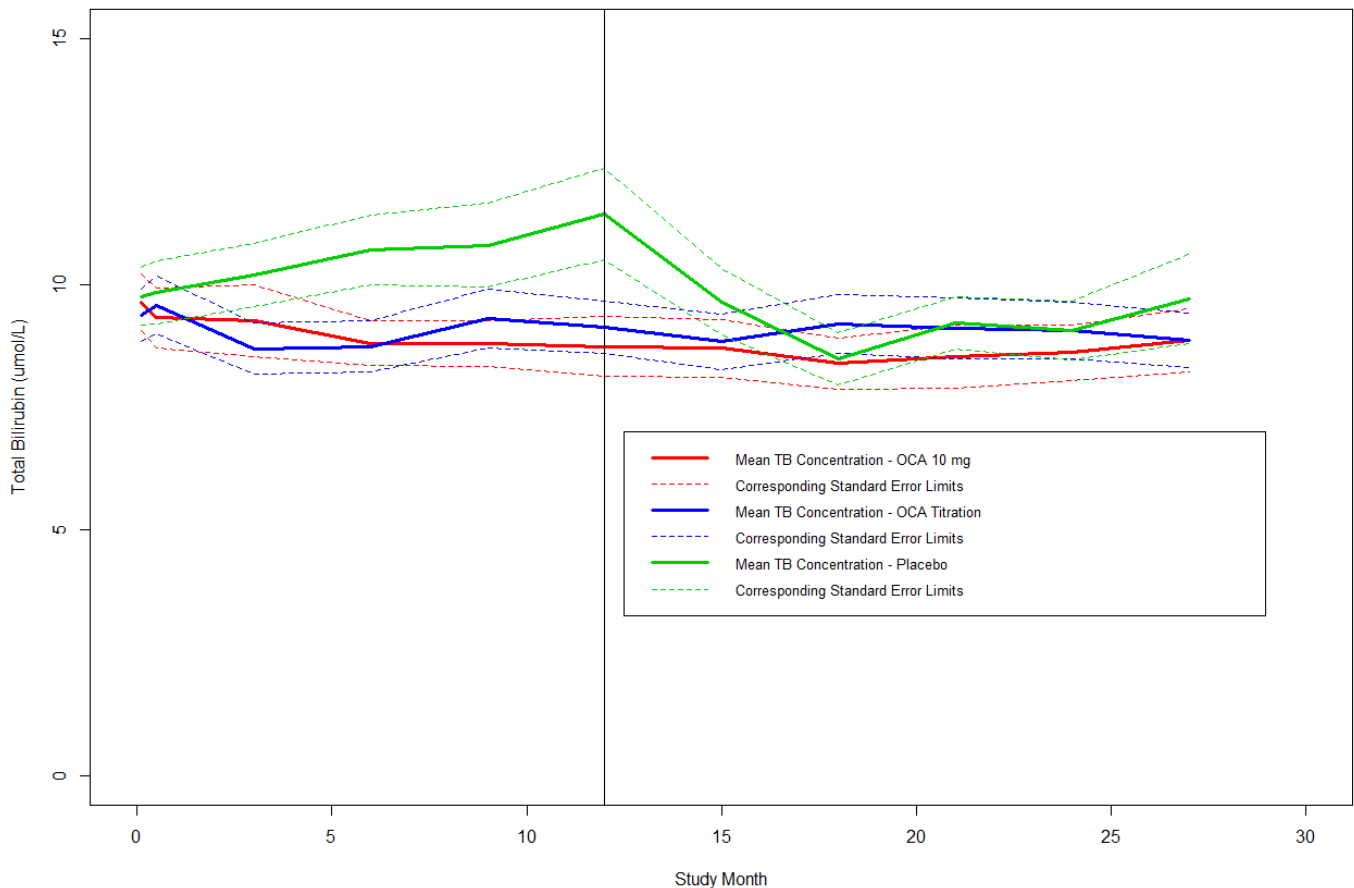
Source: Reviewer's Figure generated from integrated ADLB and ADLBOLDB datasets.

Similar to the whole ITT group results as previously presented in Figure 4, it can be seen that reductions from baseline in TB were only marginally greater in both OCA treatment groups relative to the placebo group. It can also be seen that TB levels for placebo patients slightly increased during the first 12 months; however, these levels started decreasing immediately, and ultimately remained stable, during the LTSE period once these patients started OCA administration.

As a whole in this trial, the changes in TB levels were observed to be miniscule, and TB levels were generally stable throughout OCA treatment exposure. This may very well have been attributed to the enrolled trial population primarily consisting of early stage PBC patients who were being administered UDCA. Figure 7 below displays mean TB levels over time for all

35

available subjects from the Global PBC Study who had an early stage diagnosis of PBC and who were being administered UDCA.

Figure 7
TB Concentration (μmol/L) in Early Stage PBC Subjects with UDCA Use from the Global PBC Study

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Upper Limit	6.1293	5.751	5.3649	5.3921	5.524	5.4777	4.9229	8.7731	5.4975	5.2268	7.2854	4.6034	5.1189	6.716	5.5455	6.2133
Total bilirubin	5.8635	5.4638	5.039	5.0011	5.1199	4.9321	4.3659	6.7579	4.7702	4.5342	5.945	4.1055	4.4481	5.6178	4.7988	5.168
Lower Limit	5.5977	5.1766	4.7132	4.6101	4.7157	4.3864	3.8089	4.7428	4.043	3.8417	4.6045	3.6076	3.7773	4.5195	4.052	4.1226
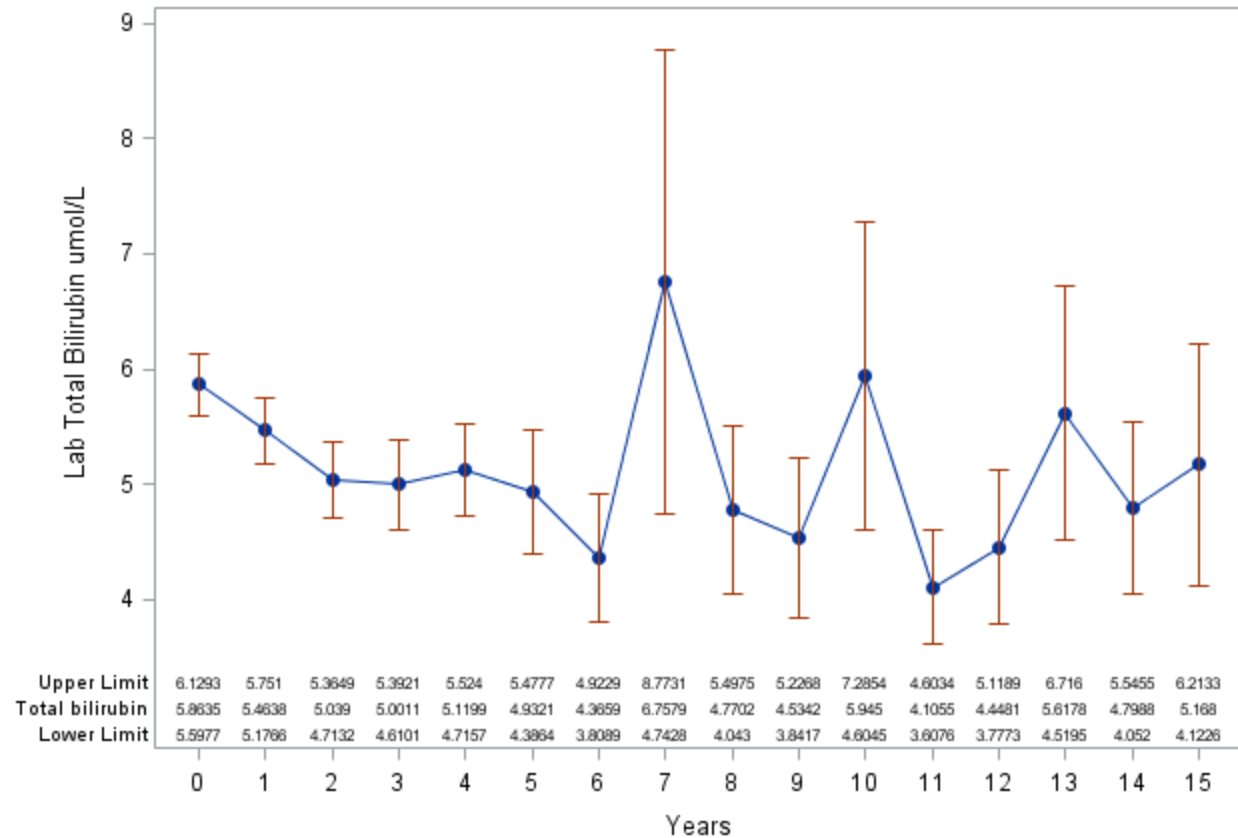
Source: Figure by Dr. Min Min generated from GPBClab_FDA dataset.

It can readily be seen from Figure 7 that these subjects exhibited reasonably flat TB levels over time. Consequently, this may render, as questionable, any potential claim that OCA therapy maintains low and stable TB levels within this patient population (i.e., early stage PBC patients using UDCA) because it appears that these TB levels would have stayed low and stable regardless of OCA intervention.

3.3 Evaluation of Safety

This evaluation is beyond the scope of this review. Please see Section 7 of the clinical review document for full details regarding the safety profile of OCA in patients with PBC.

36

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age and Geographical Region

As can be observed from Table 3 above, the overwhelming majority of study participants were female (i.e., 90.7%) and white (i.e., 94.0%); hence, this precluded any meaningful/informative subgroup analysis for gender and race. In addition, due to the limited number of patients from Australia, the subgroup analysis for geographical region was divided into two categories: Europe and North America/Australia. All subgroup analyses presented in Table 12 below pertain to the applicant's primary composite endpoint at Month 12 and are descriptive in nature.

Table 12
Proportion of Patients who Achieved Response at Month 12 – By Age and Geographical Region Subgroups
(ITT)

Statistics	10 mg OCA (N = 73)	OCA Titration (N = 70)	Placebo (N = 73)
Age Subgroups			
Age < 65	n=56	n=60	n=60
Response at Month 12 – n (%) [1] [2]	29 (51.8%)	28 (46.7%)	7 (11.7%)
Age ≥ 65	n=17	n=10	n=13
Response at Month 12 – n (%) [1] [2]	5 (29.4%)	4 (40.0%)	0
Geographical Region Subgroups			
Europe	n=51	n=45	n=49
Response at Month 12 – n (%) [1] [2]	23 (45.1%)	23 (51.1%)	3 (6.1%)
North America/Australia	n=22	n=25	n=24
Response at Month 12 – n (%) [1] [2]	11 (50.0%)	9 (36.0%)	4 (16.7%)

Source: Reviewer's Table generated from ADLIVER dataset.
Note: Denominators for percentages are n, the number of patients within each subgroup category.
[1]: A patient was designated as a responder if all three of the following conditions were met: (1) 12-Month value of ALP < 1.67×ULN; (2) 12-Month value of TB ≤ ULN; (3) ALP reduction from baseline at Month 12 ≥ 15%.
[2]: Patients with missing data at these timepoints were designated as non-responders.

It appeared that more patients who were less than 65 years of age achieved a response in all treatment groups (i.e., 10 mg OCA, OCA Titration, and Placebo) relative to those that were greater than or equal to 65 years of age (i.e., the geriatric age group). However, since only 18.5% of all ITT patients were in the geriatric age group, these findings may not be reliable. Regarding the two different geographical regions, the results appeared consistent; additionally, these results were consistent with the overall ITT population results as previously seen in Table 4 above.

37

4.2 Other Special Subgroup Populations

As can be observed from Table 3 above, the overwhelming majority of study participants were using UDCA (i.e., 92.6%) and could be designated as having early stage PBC disease (i.e., 90.3%); hence, this precluded any meaningful/informative subgroup analysis for UDCA usage and PBC disease stage. Results for the more refined special subgroup population of clinical interest pertaining to who was primarily enrolled in the 747-301 trial, i.e., early stage PBC patients who also had screening/baseline ALP ≥ 1.67×ULN and who were being administered UDCA, are already presented in Table 11 above.

5   SUMMARY AND CONCLUSIONS

5.1   Statistical Issues

Overall, the design of the 747-301 pivotal study was deemed adequate from a statistical perspective, and the applicant's corresponding SAP was deemed appropriate. There were no statistical review issues identified for this pivotal trial that would preclude product approval. Although the design, statistical analyses and results of this trial appeared to be convincing and robust, the fundamental issue of this trial, and the NDA overall, was that the patients enrolled in this phase 3 study were not adequately comparable to the broad spectrum of PBC disease patients studied by the Global PBC Group. This rendered, as questionable, the overall adequacy/applicability of the pivotal trial's primary composite endpoint, which was to be used by the applicant as a basis for accelerated approval of this NDA. In particular, the primary composite endpoint was constructed based on the overall Global PBC study results and accordingly incorporated 12 month changes/reductions in both ALP and TB levels assuming elevated levels for each parameter. However, the enrolled trial patients primarily represented the early stage PBC disease population (whose patients only exhibit elevated ALP levels as specified by the Rotterdam PBC disease staging criteria) who were also concomitantly using UDCA.

Dr. Min, an independent statistical reviewer, who was intentionally asked not to study any 747-301 trial data, conducted her review using the submitted subject-level Global PBC Study data to adequately match a clinically meaningful subset of Global PBC Study registry patients with the aforementioned majority of enrolled patients in study 747-301, while subsequently assessing whether a 12-month reduction in ALP levels alone could be reasonably likely to predict clinical outcome (i.e., death or liver transplant) in this PBC disease subpopulation. She ultimately confirmed the reasonable predictability of ALP, and the statistical team proposed a stratified cut point to further confirm OCA's efficacy in the treatment of PBC trial patients.

5.2   Collective Evidence

The applicant submitted the results from the 747-301 trial to support the efficacy of OCA for the treatment of PBC. This study was adjudicated as being adequate and well-controlled from a design and statistical analysis perspective, and the trial results showed a significant difference in the number of responders at 12 months, pertaining to the applicant's pre-specified primary composite endpoint, between both OCA treatment groups and placebo. The trial results also

38

showed that OCA substantially reduced ALP levels, relative to placebo, after 12 months. The currently ongoing open-label long term safety extension period suggests a sustained/durable OCA efficacy profile with respect to ALP levels. As a whole in this trial, the changes in TB levels were observed to be miniscule, and TB levels were generally stable throughout OCA treatment exposure. This may very well have been attributed to the enrolled trial population primarily consisting of early stage PBC patients who were being administered UDCA; these patients are understood to exhibit reasonably low and stable TB levels over time. Consequently, this may render, as questionable, any potential claim that OCA therapy maintains low and stable TB levels within this studied patient population (i.e., early stage PBC patients using UDCA) because these TB levels most likely would have stayed low and stable regardless of OCA intervention.

5.3   Conclusions and Recommendations

There is sufficient evidence in supporting the efficacy claims for OCA in the treatment of PBC. The claims reflected within the applicant's product labeling are supported by the results presented in this review. It is recommended that inferential statistics (i.e., p-values) be presented within the final product labeling for only the pairwise comparisons between the individual OCA treatment groups and placebo from the 747-301 trial in regards to the trial's pre-specified primary composite endpoint assessed at Month 12; these were the only analyses that formally controlled the overall study-wise type I error rate in this trial. All other 747-301 analysis results should be presented in a descriptive manner. (b) (4) Finally, it is recommended that confirmatory clinical outcomes study 747-302 enroll the appropriate PBC patient population to ultimately verify 12-month ALP reduction alone as a surrogate reasonably likely to predict clinical benefit.

6   APPENDIX

I would like to extend a special thanks to Dr. Min Min, Dr. Andrejus Parfionovas, Dr. Yeh-Fong Chen, Dr. Sue-Jane Wang, Dr. Ruby Mehta, Dr. Lara Dimick-Santos, Dr. Dragos Roman, and Dr. Amy Egan for making this review possible.

39

Reference ID: 3921595

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 207999
Drug Name:	OCALIVA ^{TM} (obeticholic acid; OCA)
Indication(s):	Treatment of Primary Biliary Cirrhosis (PBC) in combination with ursodeoxycholic acid (UDCA) in adults with an inadequate response to UDCA or as monotherapy in adults unable to tolerate UDCA
Applicant:	Intercept Pharmaceuticals, Inc.
Date(s):	Stamp date: 06/29/2015
	PDUFA Goal date: 05/29/2016
Review Priority:	Priority
Statistical Reviewer:	Min Min, Ph.D., DB3
Concurring Reviewers:	Yeh-Fong Chen, Ph.D., DB3
	Stephen Wilson, Dr. PH, DB3
	Sue-Jane Wang, Ph.D., OB
Medical Division:	Division of Gastrointestinal and Inborn Error Products (DGIEP)
Clinical Team:	Medical Officer: Ruby Mehta, M.D.
	Medical Team Leader: Lara L. Dimick, M.D.
Project Manager:	Anissa Davis-Williams, RN, B.S.N., M.P.H., C.P.H.M.

Table of Contents

1 EXECUTIVE SUMMARY

The sponsor submitted two efficacy trials (i.e., Phase 2 Study 747-202 and Phase 3 Study 747-301, respectively) to support the accelerated approval for obeticholic acid (OCA) in treating adult patients with primary biliary cirrhosis (PBC). To date, the only drug therapy approved for PBC is ursodeoxycholic acid (UDCA).

Following FDA's advice, the sponsor collaborated with the Global PBC Study Group to investigate whether any liver-related biochemical variables, particularly for the endpoints used in the Phase 3 Study 747-301, i.e., alkaline phosphatase (ALP) and total bilirubin (TB), could be surrogates that would reasonably likely predict clinical benefit. Overall, the analyses of the Global PBC Study Group database supported the proposition that TB and ALP at 12 months and other time points after study enrollment are significant predictors for transplant-free survival in patients with PBC. The sponsor subsequently leveraged this finding to support Subpart H approval based on the primary composite endpoint of ALP <1.67x ULN, total bilirubin ≤ULN, and ALP decrease of ≥15% from baseline at Month 12 for Phase 3 Study 747-301.

During the NDA review, we noted that Study 747-301 primarily enrolled the early disease stage PBC patients, whose baseline ALPs were at least 1.67xULN and TB measurements were within the normal range (92% of patients enrolled). However, patients in the overall Global PBC database had a much broader disease spectrum than those included in Study 747-301. It remains unclear as to whether a patients' ALP at 12 months alone can reasonably likely predict clinical outcome (i.e., death or liver transplant) in the patient population studied in Study 747-301. In addition, even if it could be used for this purpose, it appeared that it was difficult to clearly pre-specify a suitable cutoff. Therefore, we analyzed the PBC data by sub-setting patients with similar clinical demographics as those in Study 747-301 to better understand if evidence existed to support the use of ALP alone at 12 months to predict clinical outcomes for an early stage clinical population.

After sub-setting patients with normal TB at enrollment, we obtained 909 patients with 131 events from the Global PBC Study Group data for our analyses. Recall that in the original Global PBC study, there were 4845 patients with 1118 events of liver transplantation or death. In order to increase the reliability and generalizability, we randomly divided 909 patients into three small groups; (1) 25% of the data was used for model selection (2) 50% of the data was served as the training set and (3) the rest 25% of the data was used as testing set. We conducted the analyses for seventeen cutoffs and 5 covariates to select the best fit model(s) and suitable cutoff(s).

After thorough evaluation, the model with the factors of the __age__ and __baseline ALP raw lab values and ALP at 12 months__ had been chosen as the best predictive performance for death or liver transplantation based on the smallest point estimate of the AIC value. The distribution of ALP at time 0 or at 12 month is skewed. We have performed the model diagnosis and explored log transformation of ALP. We found that ALP at 12 months is an important predictive factor in the subset of subjects whose baseline ALP is at least 1.67xULN. Although the distributions of all ALP measurements (e.g., ALP and ALP lab raw values) are not perfectly symmetric, the same

3

model was chosen based on log transformation. To be consistent with Lammer's paper, we presented results based on the original scale in this review.

Study 747-301 used combination cutoff which is ALP at Month 12 less than 1.67xULN and at least 15% decrease from baseline (we call it protocol defined cutoff in this review). As one inclusion criterion of Study 747-301 was baseline ALP at least 1.67xULN, patients whose baseline ALPs between 1.67 and any other derived ALP value (>1.67xULN) can only be responders based on the percent reduction criterion. In other words, any other absolute derived ALP value (>1.67xULN) will restrict some subset of patients who become responders only based on the percent change of the ALP. According to the results shown in Table 5.4 and 5.5 of the Appendix, the combination of 2.0xULN and either 15% or 40% reduction performed better than 1.67xULN and 15% reduction, we propose the following stratified cutoff to take into account the aforementioned patients in our cutoff selections:

 (1) ALP less than 1.67xULN at Month 12 and at least 15% decrease from baseline for the patients whose baseline ALP were between 1.67 and 2.0xULN; or
(2) ALP less than 2.0xULN at Month 12 and at least 40% decrease from baseline for the patients whose baseline ALP were at least 2.0xULN)

From the above definition, our proposed stratified cutoff resulted in similar point estimates of C-statistic compared to other combined cutoffs of (a) 2.0xULN and 15%, (b) 2.0xULN and 40%, (c) 1.67+2.0xULN and 15%, (d) 1.67+2.0xULN and 40% (i.e., 0.68 to 0.69 in the training sets and 0.68 to 0.70 in the testing sets, respectively).We examined the robustness of our proposed stratified cutoff's predictability in comparison with protocol defined cutoff (i.e., ALP <1.67 ULN and 15% reduction) through subgroup analyses, including those by age, age at diagnosis, year of diagnosis, region and baseline ALP raw lab values. We found that the point estimates (hazard ratios) of the association are between the cutoffs and the clinical outcome appeared to be consistent even though some of the 95% confidence intervals were narrower or wider than those in Global PBC Study, which can be mainly due to the smaller size of the subgroups.

AC meeting will be held on April 7, 2016 to discuss our proposed stratified cutoff.

2   BACKGROUND

Primary biliary cirrhosis (PBC) is a rare, serious, and life-threatening hepatic condition primarily affecting females. PBC is characterized by cholestasis with progressive impairment of bile flow in the liver that results in increased hepatocellular bile acid concentrations. Bile acids are natural detergents, and abnormally elevated hepatocellular concentrations can be toxic to the liver. Such hepatocellular injury results in a local inflammatory response and is signaled early on by the secretion of alkaline phosphatase (ALP). The only drug therapy currently approved for PBC is ursodeoxycholic acid (UDCA). While UDCA therapy has a marked impact on clinical outcomes in PBC, up to 40% of UDCA-treated patients have a suboptimal or absent response to UDCA and as such are at significantly increased risk of an adverse outcome.

4

Due to the rarity of PBC and its slow natural history, it is very difficult to conduct trials assessing clinical endpoints and hence the FDA has provided feedback regarding the possibility of pursuing a Subpart H accelerated approval for OCA in the treatment of PBC. Based on this advice, the Sponsor helped establish and collaborated with the Global PBC Study Group project to investigate if biochemical variables, in particular ALP and bilirubin, could be used as acceptable surrogate endpoints "reasonably likely to predict clinical benefit" to support a Subpart H Accelerated Approval. With the positive findings from the aforementioned Global PBC project[1], the sponsor conducted one pivotal phase 3 Study 747-301 using the primary endpoint of ALP <1.67x ULN, total bilirubin ≤ULN, and ALP decrease of ≥15% from baseline at Month 12 for the accelerated approval for obeticholic acid (OCA) as the treatment of primary biliary cirrhosis (PBC) in adult patients. Of note, the PBC group's principle investigators are located at the Erasmus MC University Medical Center in Rotterdam, Netherlands and it was a multinational multicenter registry study that followed nearly 5,000 adult PBC patients until they achieved clinical outcome (i.e., death or liver transplant).

During the NDA review, we noted the enrolled trial patients only represented the early disease stage PBC population; this population typically exhibits elevated ALP only and TB is within the normal range. The enrolled trial population was not directly comparable to the entire Global PBC Group (see Table 2.1). Therefore, we proposed sub-setting the Global PBC Group in order to address our main concern, which is whether ALP at 12 months is predictive of clinical outcome (i.e., death or liver transplant).

Table 2.1: Baseline patient characteristics for Global PBC data and Study 747-301

	Global PBC data (N=4845)	Study 747-301 (N=216)
Age at entry (year)	54.5±12.0	55.8±10.5
Female	4348 (90%)	196 (91%)
AMA positive	4280 (88%)	194 (90%)
Year of diagnosis	1959-2012	1980-2012
Early disease stage	**2040 (42%)**	**198 (92%)**
Moderately advanced disease stage	**989 (15%)**	**18 (8%)**
Advanced disease stage	**259 (5%)**	**0 (0%)**
Bilirubin (>ULN)	**974 (26%)**	**18 (8%)**
ALP (×ULN)	2.10 (1.31-3.72)	2.40 (1.22-6.86)

Source: Sponsor's 747-301-report-body.pdf and Lammers et al. 2014 paper's Table 1.

---

[1] Lammers, W. J., et. al. Levels of Alkaline Phosphatase and Bilirubin are surrogate end points of outcomes of patients with PBC: an international follow-up study. Gastroenterology 2014:1-12

5

3 SOURCES OF DATA AND LIMITATIONS

3.1 Global PBC Data Submissions and Information Requests (IRs)

Due to the complexity of this NDA and the need of re-analyzing the Global PBC group data, we have sent five information requests (IR) to the sponsor. The detailed information for each IR and the sponsor's response are summarized in the Table 3.1.

Table 3.1 History of FDA Information Request during NDA Review

Date	Activity
8/31/2015	IR (due on 9/19/2015): requesting datasets (raw and derived) for Global PBC study and analysis programs (either by R or SAS) along with thorough data definition file(s).
9/14/2015	IR (received on 9/25/2015 and 10/25/2015): Please provide rationale and data analyses to support your selection of the cut-off for the primary surrogate endpoint, i.e., 1.67 × ULN for alkaline phosphatase, from the data of the PBC study group such as C-statistics, Youden's Index and AUCROCs \\CDSESUB1\evsprod\NDA207999\0029
11/03/2015	Global PBC data submitted through DMF: \\cdsesub4\NONECTD\MF029942\5934275
11/19/2015	IR: regarding many data issues and missing information
12/15/2016	Data resubmitted through DMF: \\cdsesub4\NONECTD\MF029942\5974175
1/07/2016	IR: requesting lab data and death/liver tranplantation date
1/28/2016	Lab data submitted through DMF: \\cdsesub4\NONECTD\MF029942\5991425
1/29/2016	IR: sharing FDA analysis plan
2/29/2016	IR response submitted by the sponsor: \\CDSESUB1\evsprod\NDA207999\0053

3.2 Limitations of Global PBC Data

To evaluate the use of the ALP and identify the best cutoff, we met and negotiated with the Global PBC group's statistician and the sponsor regarding the submission of the PBC study data. Even though we have thoroughly examined and tried our best to analyze the submitted data, we found that their data have the following limitations: (1) only the "years" of all the important dates (e.g., date of first visit, date of birth, UDCA date of start therapy, date of diagnosis of PBC, date of decompensation and end of follow-up date) were provided. (2) region information was only categorized as USA, Canada and Europe not as exact countries. (3) Global PBC database composed of an observation and retrospective registratry data, therefore a lot of

6

data were missing without any imputations. For the comparable subset, we have 7.92% (72 out of 909) missing ALP values (raw and derived) at 12 months (4) lab data were collected locally without centralization.

4   STATISTICAL EVALUATION

Our model selection was performed based on cross-validation prediction errors and Akaike information criterion (AIC) for 25% of the comparable subset of Global PBC data, optimal cutoff (s) based on C-statistics and hazard ratios for the training set which has 50% of the data and the analysis set for the rest 25% of the data. The subgroups analyses were conducted to explore the robustness of the chosen optimal cutoff(s) for different region, age, age at diagnosis, baseline ALP and diagnosis year groups. Kaplan Meier curves were used to demonstrate the predictive ability of the chosen optimal cutoff(s).

4.1   Model Selection

Data considerations

All but one patient in Study 747-301 have baseline ALP $\geq$ 1.67xULN; however, 92% of patients have normal TB, thus are in early stage PBC. Therefore, the medical review team determined that the analyses conducted based on a subset of Global PBC data   with comparable clinical demographics to those in Study 747-301 (see Table 5.1 in the Appendix) would be more applicable. In other words, it is necessary to re-analyze the PBC data by limiting to patients with baseline ALP $\geq$ 1.67xULN and normal TB thus are in the early stage (SG_DSRDAM=1) with UDCA use (UDCA=1). This sub-setting resulted in 909 patients with 131 events, compared to 4845 patients with 1118 events of liver transplantation or death conducted by Global PBC group. Patients in the PBC subset had a much lower event rate of 14% compared to the event rate of 23% in the entire Global PBC study. This finding is in line with clinical expectations given the course of disease. For this model selection, 25% (227 with 29 events) of 909 patients were randomly selected.

Candidate models

PBC is a female dominant disease, hence only age, year of diagnosis, ALP at baseline (raw or derived), duration of PBC and region were explored in the candidate models as covariates. Of note, we found that age and age at diagnosis was highly correlated and thus the age at diagnosis is not considered. Table 4.1 displays all different types of ALP at Month 12 which we considered. Note that in our analyses, the models including percentage changes from baseline were all adjusted for baseline ALP raw lab values and the absolute changes were all calculated based on the already derived data after they were converted to the ULN.

7

Table 4.1: Candidate models based on different types of measurements for ALP and covariates

ALP at 12 months	Covariates
Percentage change from baseline based on ALP lab raw values	Total 47 models: Age, diagnosis year and duration of PBC, region, ALP raw values at baseline
Absolute ALP	Total 61 models: Age, diagnosis year, ALP at baseline, region and duration of PBC

Best fit models based on cross-validations and AIC values

For the Cross validation (CV), we used 5-fold method. Our analyses were implemented through the R package "pec".

The statistical reviewer searched for the best fit models through candidate models using covariates age, baseline ALP and diagnosis year, region and duration of PBC. Based on the cross validation prediction errors and AIC (see Tables 5.2 and 5.3 in the Appendix), the best fit model included terms of age and raw ALP values at baseline in addition to the percentage change from baseline. Here we only listed results for the models including ALP at 12 months and ALP at baseline (raw or derived values). Although the distributions of all ALP measurements (e.g., ALP and ALP lab raw values) are not perfectly symmetric, the model with the factors of the **age** and **baseline ALP raw lab values and ALP at 12 months** was also chosen based on log transformation.

4.2 Exploration of Potential Cutoff(s)

About 75% (682 with 102 events) of 909 patients were randomly selected for searching the optimal cutoff(s) by using ten random splits (training vs. testing is 2:1) and 5-fold cross validation methods. Both methods give the largest C-statistic and hazard ratio to our proposed stratified cut-off, i.e., 1.67xULN and 15%, or 2.0xULN and 40%.

Ten random splits

Total of 682 patients were randomly split into two parts ten times: training set (455) and testing set (227). C-statistics and hazard ratios were calculated for each random split for both testing and training sets. We found that our proposed stratified cutoff: (1.67xULN and 15% decrease), or (2.0xULN and 40% decrease) resulted in larger point estimate of C-statistics and hazard ratios than the protocol defined cutoff. (see Tables 4.2 below and Table 5.4 in Appendix).

Table 4.2: Summary of C-statistics and hazard ratios (10 random splits)

Cut offs	C-statistic (mean)	Hazard ratio (mean)	Hazard ratio 95% CI (mean)	#significant p-values
10 Training sets				
1.67xULN and 15%	0.6395	1.82	(1.06, 3.13)	7/10
1.67xULN and 15% or 2.0xULN and 40%	0.6884	2.29	(1.33, 3.97)	10/10
10 Testing sets				
1.67xULN and 15%	0.6844	2.42	(1.08, 5.51)	4/10
1.67xULN and 15% or 2.0xULN and 40%	0.7000	2.54	(1.15, 5.69)	8/10

8

5-fold

Total of 682 patients were randomly split into 5 mutually exclusive subsets of approximately the same size of 135 patients per subset. After the first four subsets were combined, C-statistics and hazard ratios were calculated and then compared with the results in the fifth subset. The entire process was performed five times to allow each combination of 4 subsets to be pooled to serve as the training set and each subset not used in the training set to serve as the testing set. We also observed that our proposed stratified cutoff (1.67xULN and 15% decrease or 2.0xULN and 40% decrease) appears to better predict patient outcomes numerically based on C-statistics and hazard ratios (see Tables 4.3 below and Table 5.5 in Appendix).

Table 4.3: Summary of C-statistics and hazard ratios (5-fold)

	C-statistic (mean)	Hazard ratio (mean)	Hazard ratio 95% CI (mean)	#significant p-values
5 Training sets				
1.67xULN and 15%	0.6531	1.95	(1.19, 3.21)	4/5
1.67xULN and 15% or 2.0xULN and 40%	0.6924	2.32	(1.42, 3.80)	5/5
5 Testing sets				
1.67xULN and 15%	0.6775	2.38	(0.78, 7.44)	2/5
1.67xULN and 15% or 2.0xULN and 40%	0.6849	2.68	(0.89, 7.21)	1/5

4.3   Subgroup Analyses

To assess the consistency and robustness, subgroup analyses for two cutoffs (i.e., <u>our proposed stratified cutoff:</u> 1.67xULN and 15% or 2.0xULN and 40% decrease from baseline for ALP at Month 12 and <u>the protocol defined cutoff</u>: 1.67xULN and 15% reduction) were conducted and displayed in Table 5.6 in the Appendix based on the total 909 patients. We used the best fitted model including age, raw ALP at baseline and percentage change from baseline for ALP at Month 12 to perform the subgroup analyses.

We obtained similar results between our proposed stratified cutoff and the protocol defined cutoff for Study 747-301 except the diagnosis year <1990. Due to the insufficient study duration and thus only 9 events observed for patients diagnosed after Year 2000, we only considered two subsets for patients' diagnosis year (i.e., <1990 & 1990-2009). Although some of the 95% confidence intervals were narrower or wider, it can be mainly due to the smaller size of the subgroups. Figure 3 and Figure 4 display the Kaplan Meier survival curves for the protocol defined cutoff and our proposed stratified cutoff, respectively. It appears that the two curves in Figure 4 have a slight bigger separation after 10 years.

9

4.4   Forest and Kaplan-Meier Plots

Figure 1: Forest plot for subgroup analyses
Cutoff: 1.67xULN and 15% decrease from baseline for ALP at 12 months
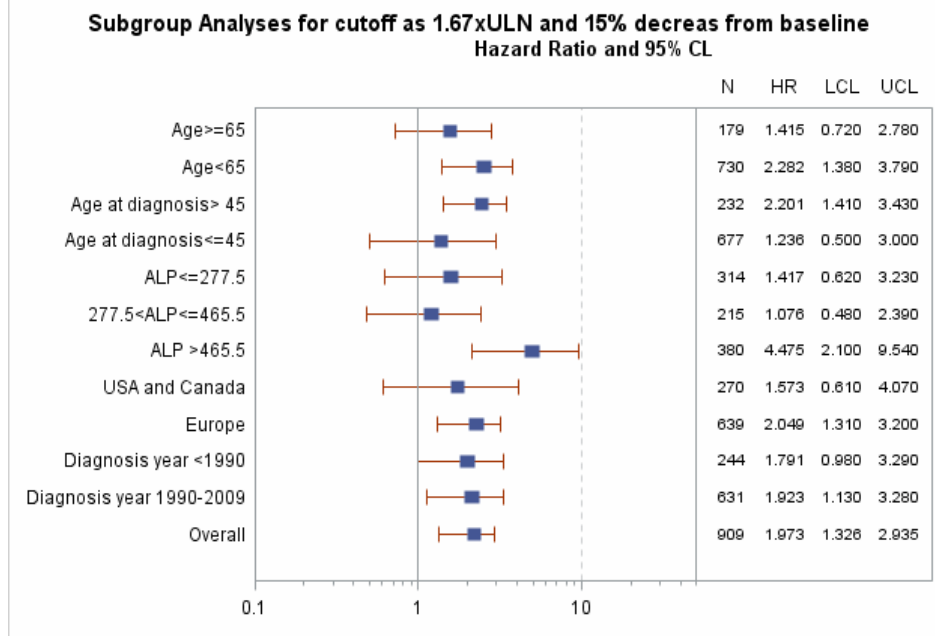
Figure 2: Forest plot for subgroup analyses
Cutoff: 1.67xULN and 15% or 2.0xULN and 40% decrease from baseline for ALP at 12 months
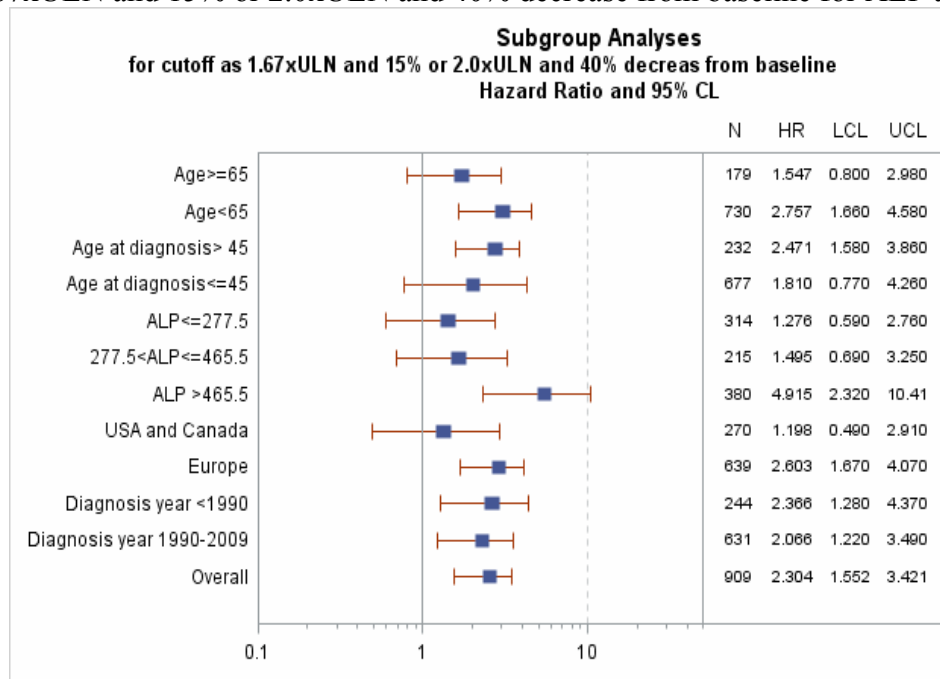
Figure 3: Kaplan-Meier plot for transplant-free survival probability
Cutoff: 1.67xULN and 15% decrease from baseline for ALP at 12 months
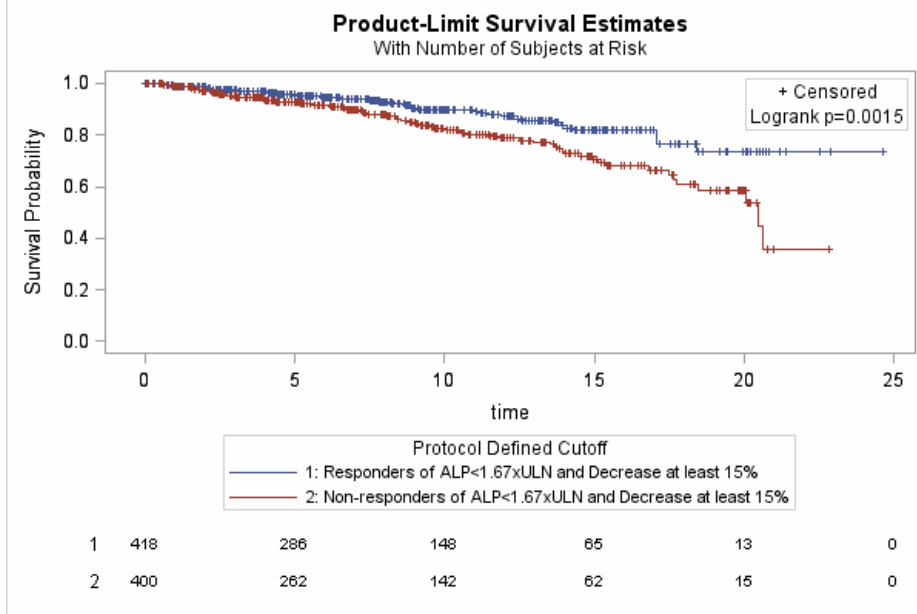
Figure 4: Kaplan-Meier plot for transplant-free survival probability
Cutoff: 1.67xULN and 15% or 2.0xULN and 40% decrease from baseline for ALP at 12 months

11
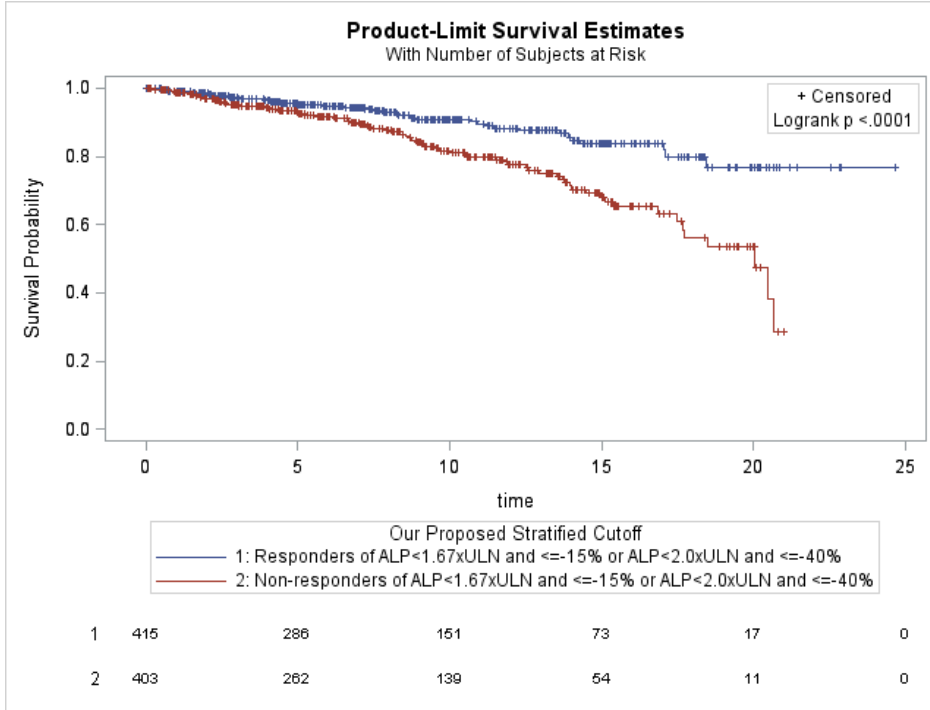
4.5   Summary of Reviewer's Findings

After sub-setting patients with similar clinical demographic as those in Study 747-301 (i.e. that had a normal TB at trial entry) we included a subset of 909 patients with 131 events in our analysis, compared to 4845 patients with 1118 events of liver transplantation or death in Global PBC database. Of these 909 patients, we randomly divided them into three parts: 25% for model selection, 50% for training set and the rest 25% as testing set. Seventeen cutoffs and 5 covariates were considered.

Our first step was to select the best fit model for each of two types of ALP at Month 12, among all the candidate models (61 models for the absolute ALP and 47 models for percentage change from baseline) by AIC and cross validation prediction error. The chosen models include the age and baseline ALP raw values when percentage from baseline was used.

Our second step was to select the optimal cutoff(s) based on C-statistics and hazard ratios by using the ten random splits and 5-fold methods. A fitted survival model including age, baseline ALP raw values and ALP at 12 months had the best predictive performance of death or liver transplantation (based on the point estimates of C-statistics and hazard ratios for time-to-event natural history data submitted by the Global PBC group) for the two types of patients: 1) those patients with baseline ALPs between 1.67 and 2.0xULN whose ALPs at Month 12 were larger than or equal to 1.67xULN and had less than 15% decrease from baseline; or 2) for those patients with baseline ALPs at least 2.0xULN whose ALPs at Month 12 were larger than or equal to 2.0xULN and had less than 40% decrease from baseline.

Our final step is to check the robustness of the predictive ability of two cutoffs by subgroup analyses. The subgroups we explored include age, age at diagnosis year, region, diagnosis year and baseline ALP raw values.

Our proposed stratified cutoff resulted in similar point estimates of C-statistic compared to other combined cutoffs of (a) 2.0xULN and 15%, (b) 2.0xULN and 40%, (c) 1.67+2.0xULN and 15%, (d) 1.67+2.0xULN and 40% (i.e., 0.68 to 0.69 in the training sets and 0.68 to 0.70 in the testing sets, respectively). To allow the responder definition captures improvement of those subjects with baseline ALP between 1.67xULN and 2.0xULN as well as those with at least 2.0xULN, our proposed stratified cutoff appears more reasonable. Furthermore, this stratified cutoff had demonstrated numerically better performance than the protocol originally defined cutoff as 1.67xULN and 15% decrease from baseline in terms of point estimates of C-statistic and hazard ratios. Also the analysis results based on 10 random splits and 5-fold were similar. Our subgroup analysis results demonstrated that the point estimates (hazard ratios) of association between the cutoffs and the clinical outcome appeared to be consistent. Although some of their 95% confidence intervals were narrower or wider, it could be mainly due to the smaller size of the subgroups. The Kaplan Meier survival curves for our proposed stratified cutoff appears to have a slight bigger separation after 10 years.

12

5 APPENDX

Table 5.1: Demographic and Baseline Characteristics of the Comparable Cohorts from Study 747-301 and the Global PBC Study

	Study 747-301 (N = 181)	Global PBC Study (N = 909)
Age at Screening (years)		
n	181	909
Mean (SD)	55.5 (9.82)	54.4 (11.16)
Median	54.0	54.0
Min, Max	29, 81	24, 86
Age Category – n (%)		
< 65 years old	151 (83.4%)	730 (80.3%)
≥ 65 years old	30 (16.6%)	179 (19.7%)
PBC Diagnosis Age (years)		
n	181	909
Mean (SD)	47.1 (10.03)	52.9 (11.24)
Median	47.0	53.0
Min, Max	25, 78	23, 86
PBC Diagnosis Age Category – n (%)		
< 45 years old	72 (39.8%)	209 (23.0%)
≥ 45 years old	109 (60.2%)	700 (77.0%)
Diagnosis Year Category – n (%)		
< 1990	2 (1.1%)	244 (26.8%)
≥ 1990	179 (98.9%)	665 (73.2%)
Duration of PBC (years)		
n	181	909
Mean (SD)	8.5 (5.63)	2.2 (3.79)
Median	7.8	0.27
Min, Max	0.4, 32	0, 36
Duration of PBC Category – n (%)		
< 7.5 years	87 (48.1%)	821 (90.3%)
≥ 7.5 years	94 (51.9%)	88 (9.7%)
Gender – n (%)		
Female	165 (91.2%)	842 (92.6%)
Male	16 (8.8%)	67 (7.4%)
Race – n (%)		
Asian	2 (1.1%)	Race
Black or African American	2 (1.1%)	Not
Other	6 (3.3%)	Available
White	171 (94.5%)	
Geographical Region – n (%)		
Australia	9 (5.0%)	0
Europe	118 (65.2%)	639 (70.3%)
North America	54 (29.8%)	270 (29.7%)

Table 5.1 continued: Demographic and Baseline Characteristics of the Comparable Cohorts from Study 747-301 and the Global PBC Study

	Study 747-301 (N = 181)	Global PBC Study (N = 909)
Total Daily UDCA Dose (mg)		
n	181	687*
Mean (SD)	1091.2 (312.66)	809.5 (233.66)
Median	1000.0	750.0
Min, Max	300, 2700	250, 1500
ALP Concentration (U/L)		
n	181	909
Mean (SD)	311.3 (95.54)	478.7 (390.77)
Median	281.5	388.0
Min, Max	200, 746	1.7, 2545
ALP Concentration (×ULN)		
n	181	909
Mean (SD)	2.621 (0.8101)	3.365 (1.770)
Median	2.380	2.722
Min, Max	1.68, 6.31	1.67, 15.30
TB Concentration (μmol/L)		
n	181	909
Mean (SD)	9.6 (4.37)	7.02 (5.65)
Median	8.3	8.0
Min, Max	2, 25	0.2, 22
TB Concentration (×ULN)		
n	181	909
Mean (SD)	0.480 (0.2077)	0.579 (0.2043)
Median	0.425	0.571
Min, Max	0.08, 0.99	0.12, 1.00

Source: Reviewer's Table generated from the 747-301 ADSL and 747-301 ADLIVER datasets along with the GPBC_FDA and GPBClab_FDA datasets.
Note: Denominators for percentages are N; '*' signifies available data.

.

14

Table 5.2: AIC values and prediction errors (33 models for absolute ALP)

Models	AIC	Prediction Error (5-fold)
Age	249.604	0.119
alp12	223.134	0.119
alp0+alp12	224.718	0.119
Diag_yr+alp12	224.892	0.119
Age+alp12	**219.955**	**0.114**
Region+alp12	225.111	0.123
Disease_duration_at0+alp12	225.072	0.121
alp0+alp12+Age	221.772	0.115
alp0+alp12+Diag_yr	226.270	0.123
alp0+alp12+Region	226.630	0.125
alp0+alp12+Disease_duration_at0+alp12	226.619	0.121
Diag_yr+Age+alp12	221.845	0.116
Diag_yr+Region+alp12	226.892	0.127
Diag_yr+Disease_duration_at0+alp12	226.830	0.124
Age+Region+alp12	221.894	0.119
Age+Disease_duration_at0+alp12	221.940	0.116
Disease_duration_at0+alp12+Region	227.059	0.126
Diag_yr+Age+alp0+alp12	223.561	0.117
Diag_yr+Region+alp0+alp12	228.253	0.130
Diag_yr+Disease_duration_at0+alp0+alp12	228.113	0.125
Age+Region+alp0+alp12	223.670	0.121
Age+Disease_duration_at0+alp0+alp12	223.767	0.117
Region+Disease_duration_at0+alp0+alp12	228.554	0.128
Age+Region+Diag_yr+alp12	223.817	0.122
Age+Disease_duration_at0+alp12+Diag_yr	223.373	0.120
Disease_duration_at0+alp12+Region+Diag_yr	228.830	0.129
Disease_duration_at0+alp12+Region+Age	223.868	0.121
Diag_yr+Region+Age+alp0+alp12	225.514	0.124
Disease_duration_at0+Diag_yr+Age+alp0+alp12	224.941	0.121
Disease_duration_at0+Region+Diag_yr+alp0+alp12	225.361	0.132
Disease_duration_at0+Region+Age+alp0+alp12	225.657	0.124
Disease_duration_at0+Diag_yr+Age+alp0+alp12	224.941	0.121
Disease_duration_at0+Regio+Diag_yr+alp0+alp12+Age	226.915	0.129

15

Table 5.3: AIC values and prediction errors (17 models for percentage change for ALP)

Models	AIC	Prediction Error (5-fold)
Age+labalp0	251.017	0.123
Percent_change_alp12+labalp0	227.522	0.128
Percent_change_alp12+labalp0+Age	**224.851**	**0.122**
Percent_change_alp12+labalp0+Diag_yr	228.733	0.134
Percent_change_alp12+labalp0+Region	229.405	0.128
Percent_change_alp12+labalp0+Disease_duration_at0	229.302	0.129
Percent_change_alp12+Diag_yr+Age+labalp0	226.304	0.126
Percent_change_alp12+Diag_yr+Region+labalp0	230.755	0.136
Percent_change_alp12+Diag_yr+Disease_duration_at0+labalp0	230.614	0.136
Percent_change_alp12+Age+Region+labalp0	226.754	0.123
Percent_change_alp12+Age+Disease_duration_at0+labalp0	226.815	0.123
Percent_change_alp12+Region+Disease_duration_at0+labalp0	231.223	0.131
Percent_change_alp12+Diag_yr+Region+Age+labalp0	228.281	0.128
Percent_change_alp12+Disease_duration_at0+Diag_yr+Age+labalp0	227.749	0.129
Percent_change_alp12+Disease_duration_at0+Region+Diag_yr+labalp0	232.604	0.124
Percent_change_alp12+Disease_duration_at0+Region+Age+labalp0	228.732	0.129
Percent_change_alp12+Disease_duration_at0+Diag_yr+Age+labalp0+Region	229.736	0.131

16

Table 5.4: Summary of C-statistics and hazard ratios (10 random splits)

Cut offs	C-statistic (mean)	Hazard ratio (mean)	Hazard ratio 95% CI (mean)	#significant p-values
10 Training sets				
1.0xULN	0.5783	2.43	(0.88, 6.99)	3/10
1.67xULN	0.6515	1.92	(1.14, 3.25)	6/10
1.76xULN	0.6614	2.07	(1.22, 3.49)	7/10
2.0xULN	0.6218	2.28	(1.36, 3.82)	10/10
3.0xULN	0.6311	2.92	(1.64, 5.23)	7/10
15%	0.6557	2.20	(1.24, 3.89)	7/10
30%	0.6280	1.75	(1.04, 2.96)	5/10
40%	0.6587	2.0	(1.19, 3.34)	9/10
60%	0.6073	1.68	(0.82, 3.51)	1/10
1.67xULN and 15%	0.6395	1.82	(1.05, 3.13)	7/10
1.67xULN and 40%	0.6509	2.12	(1.18, 3.81)	9/10
2.0xULN and 15%	0.6804	2.07	(1.22, 3.52)	10/10
2.0xULN and 40%	0.6877	2.55	(1.44, 4.50)	10/10
1 67+2.0xULN and 15%	0.6761	2.03	(1.19, 3.44)	10/10
1 67+2.0xULN and 40%	0.6841	2.51	(1.43, 4.44)	10/10
1 67 and 15% or 2.0 and 40%	0.6883	2.29	(1.33, 3.97)	10/10
1 67 and 40% or 2.0 and 15%	0.6746	2.15	(1.26, 3.66)	10/10
10 Testing sets				
1.0xULN	0.6087		(0.50, 6.89)	1/10
1.67xULN	0.6840	2.57	(1.15, 5.80)	4/10
1.76xULN	0.6880	2.52	(1.16, 5.53)	4/10
2.0xULN	0.6930	2.76	(1.31, 5.83)	7/10
3.0xULN	0.6730	3.67	(1.64, 8.23)	5/10
15%	0.6914	2.31	(1.05, 5.13)	6/10
30%	0.6749	2.27	(1.09, 4.76)	5/10
40%	0.6967	2.38	(1.13, 5.00)	6/10
60%	0.6336	2.33	(0.73, 7.83)	3/10
1.67xULN and 15%	0.6844	2.42	(1.08, 5.51)	4/10
1.67xULN and 40%	0.6849	2.69	(1.09, 6.86)	6/10
2.0xULN and 15%	0.7004	2.54	(1.19, 5.46)	6/10
2.0xULN and 40%	0.6992	2.82	(1.22, 6.64)	8/10
1 67+2.0xULN and 15%	0.7007	2.48	(1.16, 5.32)	6/10
1 67+2.0xULN and 40%	0.6986	2.78	(1.20, 6.54)	8/10
1 67 and 15% or 2.0 and 40%	0.7000	2.54	(1.15, 5.69)	8/10
1 67 and 40% or 2.0 and 15%	0.7053	2.57	(1.19, 5.59)	7/10

17

Table 5.5: Summary of C-statistics and hazard ratios (5-fold)

	C-statistic (mean)	Hazard ratio (mean)	Hazard ratio 95% CI (mean)	#significant p-values
5 Training sets				
1.0xULN	0.592	2.06	(0.91, 4.68)	1/5
1.67xULN	0.663	1.98	(1.23, 3.20)	5/5
1.76xULN	0.669	2.09	(1.30, 3.36)	5/5
2.0xULN	0.641	2.34	(1.47, 3.73)	5/5
3.0xULN	0.642	3.06	(1.81, 5.18)	5/5
15%	0.6620	2.22	(1.32, 3.72)	4/5
30%	0.6413	1.88	(1.17, 3.02)	4/5
40%	0.6582	2.05	(1.29, 3.27)	5/5
60%	0.6055	1.68	(0.89, 3.17)	1/5
1.67xULN and 15%	0.6531	1.95	(1.19, 3.21)	4/5
1.67xULN and 40%	0.6526	2.18	(1.28, 3.70)	5/5
2.0xULN and 15%	0.6893	2.21	(1.36, 3.58)	5/5
2.0xULN and 40%	0.6919	2.54	(1.52, 4.22)	5/5
1 67+2.0xULN and 15%	0.6877	2.16	(1.33, 3.50)	5/5
1 67+2.0xULN and 40%	0.6917	2.50	(1.50, 4.16)	5/5
1 67 and 15% or 2.0 and 40%	0.6924	2.32	(1.42, 3.80)	5/5
1 67 and 40% or 2.0 and 15%	0.6839	2.26	(1.39, 3.66)	5/5
5 Testing sets				
1.0xULN	0.581		(0.33, 8.49)	0/5
1.67xULN	0.666	1.79	(0.82, 6.53)	0/5
1.76xULN	0.673	2.42	(0.87, 6.76)	1/5
2.0xULN	0.725	2.67	(0.99, 7.24)	1/5
3.0xULN	0.692	3.70	(1.13, 12.5)	1/5
15%	0.6796	2.68	(0.86, 8.57)	1/5
30%	0.6989	2.91	(0.92, 9.51)	1/5
40%	0.6945	2.61	(0.88, 8.02)	1/5
60%	0.6404	2.09	(0.52, 8.63)	0/5
1.67xULN and 15%	0.6775	2.38	(0.78, 7.44)	2/5
1.67xULN and 40%	0.6628	2.35	(0.77, 7.25)	0/5
2.0xULN and 15%	0.7014	2.54	(0.88, 7.42)	1/5
2.0xULN and 40%	0.6967	2.74	(0.91, 8.42)	1/5
1 67+2.0xULN and 15%	0.6989	2.49	(0.86, 7.27)	1/5
1 67+2.0xULN and 40%	0.6961	2.71	(0.90, 8.31)	1/5
1 67 and 15% or 2.0 and 40%	0.6849	2.68	(0.89, 8.26)	1/5
1 67 and 40% or 2.0 and 15%	0.6962	2.50	(0.88, 7.21)	1/5

Table 5.6: Summary of subgroup analyses for hazard ratios (HRs)

		N	1.67xULN and 15% decrease		1.67xULN and 15% decrease or 2.0xULN and 40% decrease	
			HR	95% CI	HR	95% CI
Age (years)	≥ 65	179	1.415	(0.72, 2.78)	1.547	(0.80, 2.98)
	< 65	730	2.282	(1.38, 3.79)	2.757	(1.66, 4.58)
Age at diagnosis (years)	>45	677	2.201	(1.41, 3.43)	2.471	(1.58, 3.86)
	≤45	232	1.236	(0.50, 3.0)	1.810	(0.77, 4.26)
ALP baseline raw values (u/l)	≤277.5	314	1.417	(0.62, 3.23)	1.276	(0.59, 2.76)
	>277.5 and ≤465.5	215	1.076	(0.48, 2.39)	1.495	(0.69, 3.25)
	>465.5	380	4.475	(2.10, 9.54)	4.915	(2.32, 10.41)
Region	USA and Canada	270	1.573	(0.61, 4.07)	1.198	(0.49, 2.91)
	Europe	639	2.049	(1.31, 3.20)	2.603	(1.67, 4.07)
Diagnosis year	<1990	244	1.791	**(0.98, 3.29)**	2.366	**(1.28, 4.37)**
	1990-2009	631	1.923	(1.13, 3.28)	2.066	(1.22, 3.49)

Due to small number of events in diagnosis year 2000-2009, it was merged with 1990-1999 as one category.

Reference ID: 3898759
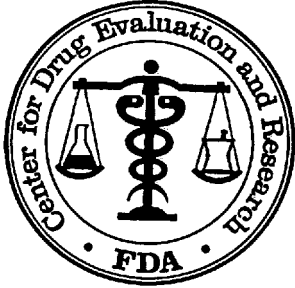
STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

NDA No.:	207999
Date Received by the Center:	Sep 22, 2015
Drug Name:	Obeticholic Acid
Indication:	(b) (4)
Sponsor:	Intercept Pharmaceuticals, Inc.
Review Finished:	Nov 5, 2015
Name of Statistical Reviewer:	Yu-Ting Weng, Ph.D.
Name of Requestor:	Ben Stevens/ Office of Pharmaceutical Quality

Primary Reviewer: Yu-Ting Weng, Mathematical Statistician, CDER/OTS/OB/DB VI
Secondary reviewer: Xiaoyu Dong, Mathematical Statistician, CDER/OTS/OB/DB VI

Concur: _____

Meiyu Shen, Ph.D., Team Leader, CDER/OTS/OB/DB VI

Distribution: NDA 207999
CDER/OPQ/ONDP/ Ben Stevens
CDER/OTS/OB/DB VI/Xiaoyu Dong
CDER/OTS/OB/DB VI/Meiyu Shen
CDER/OTS/OB/DB VI/Yi Tsong
CDER/OTS/OB/DB VI/Guoxing Soon
CDER/OTS/OB/DB VI/Lillian Patrician

Table of Contents

List of Tables

1. Executive Summary

This review provides the stability evaluation of Obeticholic Acid drug substance under ███████ (b) (4) ███████████████████████████. Assuming that the stability time trend beyond the last observed time remains the same, the conclusion of our independent stability analysis is summarized below.

- The stability data support the proposed shelf life of ██ (b)(4) months for drug substance under ███████████████████████████ (b) (4) because the 95% confidence limits of all tests are within the acceptance criteria through the proposed shelf life of ██ (b)(4) months as shown in Table 1.

Please note, the sponsor did not provide the data at the 9th month.

Table 1: Reviewer's Summarized Stability Analyses Results for Obeticholic Acid Drug Substance based on Long Term Stability Data (LCL = Lower Confidence Limit of Mean, UCL = Upper Confidence Limit of Mean)

Variable	Primary Batch Number	Last Observed Month	Prediction	95% LCL	95% UCL	Acceptance Criteria	Support a Shelf Life of (b) (4) Months?
Assay (%LC)	MT1207001	(b) (4)	(b) (4)	98.97	(b) (4)	(b) (4)	Yes
	JP1307001						
	JP1306001						
Impurity (b)(4) (%)	MT1207001			NA			Yes
	JP1307001						
	JP1306001						
Impurity total (%)	MT1207001			NA			Yes
	JP1307001						
	JP1306001						
(b) (4)	MT1207001			NA			Yes
	JP1307001						
	JP1306001						

The detailed analyses are shown in Section 3 and all stability plots are listed in the Attachment at the end.

2. Introduction and Background

On Sep 22, 2015, Office of Pharmaceutical Quality consulted the CMC statistical team in Office of Biostatistics to evaluate the stability data for Obeticholic Acid drug substance for evaluating the proposed shelf life of ██ (b)(4) months under ███████████████████████████ (b) (4) using the three primary stability lots.

OPQ CMC reviewer has one question in the consultation request.

Question 1. From the standpoint of Statistics, has this data analysis been carried out properly and does the analysis adequately support the proposed shelf life of (b)(4) months under (b)(4) [redacted] ?

On Oct 6, 2015, the Agency requested the sponsor to provide more stability data for Obeticholic Acid drug substance via Information Request Letter.

On October 28, 2015, the sponsor provided the following data:
- (b)(4) month data for 2 primary lots and (b)(4) month data for 1 primary lot under (b)(4) [redacted]
- (b)(4) month data for 3 registration lots under (b)(4) [redacted]
- (b)(4) month data for 3 registration lots under the (b)(4) [redacted]

This review will focus on the evaluation of the proposed shelf life of (b)(4) months under (b)(4) [redacted] or Obeticholic Acid drug substance.

2.1   Data Analyzed and Sources

Data are summarized in Table 2.

Table 2: Stability Data Summary

Test Variable	Batches	Testing Time Point (Months) (b)(4)
Assay (%LC), Impurity (b)(4) (%), Impurity total (% (b)(4)	MT1207001	
	JP1307001	
	JP1306001	

Based on the above data, FDA statistics reviewer conducted independent stability analyses. The conclusions are summarized in Section 1 and detailed analyses are shown in Section 3 in this review. In addition, all stability plots are provided in the Attachment at the end.

2.2   Acceptance criterion

The assay acceptance criterion is (b)(4) %LC to (b)(4) %LC.
The impurity (b)(4) acceptance criterion is $\leq$ (b)(4) %.
The impurity total acceptance criterion is $\leq$ (b)(4) %.
The (b)(4) acceptance criterion is $\leq$ (b)(4) %.

3.   Statistical reviewer's statistical analysis

This is a regular statistical evaluation of stability analyses. The data generated from three primary stability lots.

We have the following comment regarding the stability data provided by the sponsor.

In Q1A guidance, the frequency of the test period at ▓▓▓▓▓▓▓▓▓▓ (b) (4) is usually every ▓▓▓▓▓▓▓▓▓▓ (b) (4). Thus if the sponsor intends to evaluate the shelf life for the drug substance using the three primary stability batches, the sponsor should submit the stability data at 9 months. However, after the information request, the sponsor does not provide the justification for the lack of the data at 9 months.

We analyzed the stability data using the statistical approach recommended by Q1E guidance. According to the guideline, the pooling tests of slope and intercept are performed in a hierarchical order. If the data can be pooled, we estimate the shelf life at ▓(b)(4) months under the common slope or intercept model.

Table 3 lists the predicted value of assay, impurity ▓(b)(4) impurity total ▓▓▓▓▓▓ (b) (4) and their 95% confidence intervals at ▓(b) months. Stability trends of these three batches for assay, impurity ▓(b)(4) impurity total ▓▓▓▓▓▓ (b) (4) are shown in Figures 1 to 10 (Appendix).

From Table 3 and Figures 1 to 10 (Appendix), the results show that assay, all impurity attributes, ▓▓▓▓▓▓ (b) (4) months are within the acceptance criteria.

Table 3: Stability Analysis Results of All attributes at ▓(b)(4) Months for Obeticholic Acid Drug Substance (LCL = Lower Confidence Limit of Mean, UCL = Upper Confidence Limit of Mean)

Variable	Batches	Last Observed Time Point (Months)	Proposed Storage Duration (Months)	Model	Intercept	Slope	Prediction	95% LCL	95% UCL	Acceptance Criteria
Assay (%LC)	MT1207001 JP1307001 JP1306001		(b) (4)	Common intercept & slope			(b) (4)	98.97		(b) (4)
Impurity (b)(4) (%)	MT1207001 JP1307001 JP1306001			Common slope				NA		
Impurity total (%)	MT1207001 JP1307001 JP1306001			Different intercept & slope				NA		
(b) (4)	MT1207001 JP1307001 JP1306001			Common slope				NA		

4.   Conclusions and Recommendation

From Table 3 and Figures 1 to 10 (Appendix), the results show that assay, all impurity attributes, ▓▓▓▓▓▓ (b) (4) months are within the acceptance criteria. Therefore, the stability data can adequately support the proposed shelf life of ▓(b)(4) months under ▓▓▓▓▓▓▓▓ (b) (4) ▓▓▓▓.

5.   Appendix

In the appendix, we provide the stability plots for Assay, Impurity ▮(b)(4)Impurity total ▮(b)(4) ▮(b)(4) at the long-term storage conditions. Please note, in all stability plots, the dots are the observed stability data, the dashed line is the predicted mean value obtained via linear regression, and the shadow curves are the 95% confidence interval of the mean value. The straight lines are the acceptance criteria.

Figure 1 Assay versus time trend for all ▮(b)(4) batches

(b) (4)

Figure 2 Impurity (b)(4) versus time trend for Batch MT1207001

(b) (4)

Figure 3 Impurity (b)(4) versus time trend for Batch JP1307001

(b) (4)

Figure 4 Impurity (b) (4) versus time trend for Batch JP1306001

(b) (4)

Figure 5 Impurity total versus time trend for Batch MT1207001

(b) (4)

Figure 6 Impurity total versus time trend for Batch JP1307001

(b) (4)

Figure 7 Impurity total versus time trend for Batch JP1306001

(b) (4)

Figure 8 [redacted] (b) (4) versus time trend for Batch MT1207001

(b) (4)

Figure 9 [redacted] (b) (4) versus time trend for Batch JP1307001

(b) (4)

Figure 10 [redacted] (b) (4) versus time trend for Batch JP1306001

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

CARCINOGENICITY STUDIES

NDA/BLA #:	NDA 207999 ^{(b) (4)} (IND 63307)
Drug Name:	INT-747 (obeticholic acid)
Indication(s):	The treatment of primary biliary cirrhosis (PBC) ^{(b) (4)}
Applicant:	Intercept Pharmaceuticals, Inc. (Intercept)
	4350 La Jolla Village Drive, Suite 960, San Diego, USA 92122
	^{(b) (4)}
Date(s):	Received 2/25/2015; Desired Completion Date: 5/4/2015
Documents Reviewed:	Study ^{(b) (4)} 661037 (rats) and Study ^{(b) (4)} 661038 (mice) reports and electronic datasets submitted with the electronic submission on 12/19/2014 (via S-0000).
Review Priority:	Priority Review
Biometrics Division:	Division of Biometrics VI
Statistical Reviewer:	Feng Zhou, M.S.
Concurring Reviewers:	Mohammad Atiar Rahman, Ph. D.
Team Leader:	Karl Lin, Ph.D.
Medical Division:	Division of Gastroenterology and Inborn Errors Products (DGIEP)
Pharmacology Team:	Tracy Behrsing, Ph.D; Sushanta Chakder, Ph.D
Project Manager:	Anissa Davis, Senior Regulatory Health Project Manager
Keywords:	Carcinogenicity, Dose response

Table of Contents

File Name: outputfile792527110.pdf

1   Summary

This review evaluates statistically the tumorigenicity data of 2-year oral carcinogenicity studies of INT-747 in rats and mice. The review analyzes the dose-response relationship of tumor incidence and mortality (including tumor-related mortality). Tumor analyses consisted of trend analyses for dose-response relationship in tumor incidence and pairwise comparisons in tumor incidences of individual treatment groups with control. The review concludes that INT-747 decreased survival in female rats and increased tumor incidences in several organs in male and female rats. The male rats showed increases in the incidence of total hemangioma and papilloma in the paws; but the combination of all papilloma tumors in organ of ear, tail, skin, and paw did not show any statistically significant in trend analyses for dose-response and pairwise comparison between any treatment groups with controls. The female rats showed increases in the incidence of granular cell tumors in the reproductive system (cervix and vagina), benign granulosa cell tumor in ovaries, and fibrosarcoma in the skin. The drug had no effects on survival or tumor incidence in either sex in mice.

Rat Study: Rats (65/sex/dose) were dosed by oral gavage with INT-747 daily for up to 104 weeks. The INT-747 dose was 0, 0, 2, 7, or 20-mg/kg in the control 1 (C1), control 2 (C2), low-dose (LD), mid-dose (MD), and high-dose (HD) groups, respectively, in both males and females. The control groups (0 mg/kg/day) received the vehicle only.

Survival analyses showed dose-related significant decreases in survival rates in females, but not in males. The respective survival rate in C1, C2, LD, MD, and HD groups at terminal sacrifice were 58%, 60%, 55%, 42%, and 46% in females (p=0.0594) and 31%, 28%, 28%, 26%, and 25% in males. The pairwise comparisons show statistically significant decreases in survival rates in MD group compared to the pooled controls (p=0.0487) and compared to C2 group (p=0.0363) in female rats.

Tumor analyses showed positive responses in both trend analyses and pairwise comparisons in tumor incidences in both sexes. In males, the trend analysis showed statistically significant increase in the incidence of total hemangioma (both benign and malignant combined) (p= 0.0077) and benign papilloma tumor in paws (p=0.0081). The pairwise comparison showed that only the HD group had statistically significant increases in tumor incidence in these two tumor types when compared to the pooled controls (p= 0.0393, and p= 0.0352, respectively). The combination of all papilloma tumors in organ of ear, tail, skin, and paw did not show any statistically significant in analyses for dose-response and pairwise comparison between any treatment groups with controls.

In female rats, tests on tumor data showed a statistically significant positive trend (with p<0.025) in following tumor types: benign granular cell tumors in cervix (p=0.0029), in vagina (p=0.0064), and combination of the two organs (p<0.001); benign granulosa cell tumors in ovaries (p=0.0129); and malignant fibrosarcoma (p=0.0105), benign+malignant fibroma (p=0.0200) in skin, and malignant fibrosarcoma in skin/subcutis (p=0.0216). The pairwise comparisons also showed statistically significant increased incidence in HD group compared to the combined controls (with p values less than 0.05). These tumors included benign granular cell tumor in cervix (p=0.0030) and in vagina (p=0.0309), and combination of two organs (p<0.001); benign granulosa cell tumor in ovaries (p=0.0309), malignant fibrosarcoma (p=0.0377) and fibrosarcoma tumor (benign or malignant) in skin (p=0.0377), and malignant fibrosarcoma (p=0.0377) and malignant fibrosarcoma (p=0.0377) in skin/subcutis.

File Name: outputfile792527110.pdf

Mouse Study: Mice (65/sex/dose) were dosed by oral gavage with INT-747 daily for up to 104 weeks. The INT-747 dose was 0, 0, 4, 10, or 25-mkd in the C1, C2, LD, MD, and HD groups, respectively, in both sexes. The control groups (0 mg/kg/day) received the vehicle only.

The survival analysis showed that the HD females had a statistically significant decrease in mortality when compared to the pooled controls. The trend test and pairwise comparisons did not show statistically significant increased mortality in any treated group in male mice. The respective survival rates in the C1, C2, LD, MD and HD groups at the termination (Week 104) were 42%, 37%, 38%, 43%, and 63% in females and 34%, 46%, 40%, 34%, and 38% in males .

In female mice, tests on tumor data showed a statistically significant increasing trend (p=0.0119<0.025) in carcinoma, hepato in liver; the pairwise comparison show a numerically increased incidence in HD group compared to the combined controls (p=0.0584); the p-value was slightly over 0.05 alpha levels.

2   Background

The sponsor conducted 24-month carcinogenicity studies by oral (gavage) administration in CD rats and CD-1 mice (one each). This review analyzed the SAS data sets of these studies received on 12/19/2014 via submission NDA207999/S0000. The body weight, body weight change and food consumption data was analyzed by ⬛⬛⬛ (b) (4) and the statistical evaluation of survival data and tumor incidence was performed by ⬛⬛⬛ (b) (4)

The phrase "dose response relationship" refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases. The mg/kg/day will be referred to as mkd. Results of this review have been discussed with the reviewing pharmacologist Dr. Tracy Behrsing.

3   Rat Study

Study Report: ⬛ (b) (4) **661037.pdf; SAS data: 661037ft.xpt and 661037mt.xpt**

This study assessed the carcinogenic potential of INT-747 in male and female CD rats. The test material was administered at doses of 2, 7 or 20 mkd of INT-747 once daily by oral gavage. This review refers these dose groups as the low (LD), mid (MD), and high (HD) dose groups, respectively. There were two control groups; the control animals received the vehicle (0.5% carboxymethylcellulose [CMC] in deionized water). The dose volume for all groups was 10 mL/kg. There were 65 rats/sex/dose. In addition, 5 rats/sex in control group and 8 rats/sex in the treated groups served as toxicokinetic animals. The study was originally designed to continue for at least 104 weeks; however the male rats were terminated at week 101 due to the high mortality (less than 15 surviving animals in one group). The female groups were terminated during week 104 as planned.

Activities performed included clinical examinations (twice daily examinations for mortality and signs of ill health or reaction to dosing, daily post dosing cage side observations, weekly detailed examinations, palpable mass examinations on main study animals), body weights, food consumption, laboratory investigations (hematology) for health screen prior to dosing initiation and for main study animals at study termination (blood smears). Post-mortem evaluations for main study animals included macroscopic and microscopic observations.

File Name: outputfile792527110.pdf

3.1 Sponsor's Analyses

3.1.1 Survival Analysis

The sponsor calculated the Kaplan-Meier estimates (Kaplan and Meier, 1958) of group survival rates by sex and presented the graphics. The generalized Wilcoxon test (Gehan, 1965) was used to compare the homogeneity of survival rates across the groups at the 0.05 significance level. If the survival rates were significantly different, the generalized Wilcoxon test (Gehan, 1965) was used to make pair-wise comparisons of each INT-747-treated group with the control groups. A log-rank dose-response trend test of survival rates was also performed including the pooled control group and 3 INT-747-treated groups. Survival times in which the status of the animal's death, planned interim sacrifice or terminal sacrifice, were considered censored values (survival times for which the death event did not occur) for the purpose of the Kaplan-Meier estimates (Kaplan and Meier, 1958) and survival rate analyses.

Sponsor's findings: The sponsor reported the total unscheduled deaths in the following table. The survival rates ranged from 23% to 31% for males and 41% to 60% in females. There appeared a dose-related decreasing trend in the survival rates in both sexes. The sponsor's analysis showed that the decreasing trend reached the statistical significance when compared with the pooled control group (p=0.0298) and C2 group individually (p=0.0414) in female rats. There were no other statistically significant findings among males or females.

The Sponsor's Report of Total Unscheduled Deaths, n=65/sex/dose group)

Dosage (mg/kg/day)	0	0	2	7	20
Males [a]	65	65	65	65	65
Number survival	20	18	17	16	15
Number of animals found dead or euthanized in extremis	45	45	48	49	50
Accidental deaths	0	2	0	0	0
% survival [b]	31%	29%	26%	25%	23%
Females [a]	65	65	65	65	65
Number survival	38	39	35	26	29
Number of animals found dead or euthanized in extremis	25	26	30	37	34
Accidental deaths	2	0	0	2	2
% survival [b]	60%	60%	54%	41%	46%

[a] = Number of animals per dose group.
[b] = % survival was corrected based on the number of accidental deaths.

[Source: page 48 of study report of (b) (4) 661037.pdf]

3.1.2 Tumor Data Analysis

The sponsor analyzed the tumor incidence data using the Peto's mortality-prevalence method, without continuity correction, incorporating the context (incidental or fatal) in which tumors were observed. The following fixed intervals were used for incidental tumor analyses: weeks 0-50, 51-80, 81-end of study, and scheduled terminal sacrifice. For each sex, tumors that were detected, either by

File Name: outputfile792527110.pdf

palpation or necropsy, after the first animal of that sex was terminally sacrificed were considered incidental and included in the scheduled terminal sacrifice interval for analyses. For example, tumors that were detected in male rats after the first sacrifice on study day 704 were considered incidental and included in the scheduled terminal sacrifice interval for the purpose of statistical analysis. Tumors classified as mortality-independent, such as, but not limited to, those of the mammary gland and skin, were analyzed with Peto's mortality-independent method incorporating the day of detection.

Adjustment for multiple testing: In order to control the false positive error, the sponsor tested the common and the rare tumors at 0.005 and 0.025 significance levels, respectively (Lin, 2000) for positive dose response relationship, and 0.01 and 0.05 for pairwise comparisons. Tumors are considered as common with a background rate of $\geq 1\%$ and as rare with a background incidence of $< 1\%$.

Sponsor's findings:
In male rats, there was a statistically significant increase in the incidence of pheochromocytoma, malignant and the combination pheochromocytoma malignant/benign in the adrenal medulla when comparing the low dose level with the pooled control group. There was a statistically significant increasing trend in the incidence of hepatocellular adenoma in the liver when compared to the pooled controls and the C2 group. There was also a statistically significant increase in the incidence of this tumor when comparing the HD group with the pooled controls.

There was a statistically significant increasing trend in the incidence of systemic hemangioma and the combination hemangiosarcoma/hemangioma when compared to the pooled controls and the C2 group. There was also a statistically significant increase in the incidence of hemangioma and the combination hemangiosarcoma/hemangioma when comparing the HD group with the pooled controls. There were no other statistically significant tumor findings among males.

In female rats, there was a statistically significant increasing trend in the incidence of benign granular cell tumor in the cervix when compared to the pooled controls. There was a statistically significant increasing trend in the incidence of lipoma in the kidneys when compared to the pooled controls. There was a statistically significant increasing trend in the incidence of benign granulosa cell tumor in the ovaries when compared to the pooled controls, C1 and C2 groups. There was also a statistically significant increase in the incidence of the tumor when comparing the HD group with the pooled controls. In addition, there was a statistically significant increasing trend in the incidence of the combination malignant thecoma/benign granulosa cell tumor in the ovaries when compared to the C2 group. There was a statistically significant increasing trend in the incidence of fibrosarcoma in the skin when compared to the C1 group. There was a statistically significant increase in the incidence of fibrosarcoma in the combined skin/subcutis when comparing the HD group with the pooled controls.

There was a statistically significant increase in the incidence of polyp in the uterus when comparing the low dose level with the pooled controls, control 1 and control 2 groups. There was a statistically significant increasing trend in the incidence of benign granular cell tumor in the vagina when compared to the pooled controls, C1 and C2 groups. There was also a statistically significant increase in the incidence of the tumor when comparing the HD group with the pooled controls. There were no other statistically significant tumor findings among females.

The sponsor claimed that at least 21 males per group and 41 females per group survived into week 90 suggesting that the exposure to test article was adequate for an informative interpretation of the tumor incidence analysis results.

File Name: outputfile792527110.pdf

The sponsor stated that the FDA draft Guidance Document (May, 2001) addresses the interpretation of statistical significance when two identical control groups are used in a carcinogenicity study. According to that document, a reasonable approach to take is to consider a trend or difference in tumor rates as significant only if it is significant when compared to each of the control groups (at page 26). In conclusion, there was no INT-747-related effect on survival or the development of neoplastic lesions. Therefore, INT-747 was not considered carcinogenic in the rat.

Reviewer's comment: Based on the agreement with the pharmacologist, the tumor test will be based on the pooled control groups.

3.2   Reviewer's Analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing pharmacologist, this reviewer performed survival and tumor data analyses using data submitted electronically in NDA 207-999 on 12/19/2014.

3.2.1   Survival Analysis

The survival distributions of rats in all treatment groups were estimated using the Kaplan-Meier product limit method. For combined control, low, medium, and high dose groups, the dose response relationship was tested using the likelihood ratio test and the homogeneity of survival distributions was tested using the log-rank test. The Kaplan-Meier curves for survival rates are given in Figures 1A and 1B in the appendix for male and female rats, respectively. The intercurrent mortality data are given in Tables 1A and 1B in the appendix for male and female rats, respectively. Results of the tests for dose response relationship and homogeneity of survivals, are given in Tables 3A and 3B in the appendix for male and female rats, respectively.

Reviewer's findings: This reviewer's analysis showed the numbers (percent) of survival was 20 (31%), 18 (28%), 18 (28%), 17 (26%), and 16 (25%) in male rats and 38 (58%), 39 (60%), 36 (55%), 27 (42%), and 30 (46%) in female rats in the C1, C2, LD, MD, and HD groups, respectively. The tests didn't show a statistically significant dose response relationship in mortality across combined control and treated groups in male or female rats. The pairwise comparisons show statistically significant increased mortality in the MD group compared to the pooled controls (p=0.0487) and compared to the C2 (p=0.0363) in female rats.

Reviewer's comment: The sponsor's analysis showed 20(31%), 18 (29%), 17(26%), 16(25%), and 15(23%) survivors, while this reviewer's analysis showed 20 (32%), 18 (28%), 18 (28%), 17 (26%), and 16 (25%) in C1, C2, LD, MD, and HD group for male rats, respectively. These differences are due to the facts that there were five male rats (#6370, in C1, #6358 in LD, #6604 in MD, and #6414, and #6455 in HD) that died naturally during their respective terminal sacrifice week (101$^+$ week). These five male rats lived up to their respective terminal sacrifice week and the 77% of tumors found in these male rats were benign. The tumors found in the animals (#6370 and #6455) cause death. This reviewer classified these animals which the tumors did not cause the death as survivors, while the sponsor counted them as dead. Also, there were two male rats (#6564 and #6397 in C 2) that died accidently before week 28. The sponsor's analysis excluded these two animals while this reviewer's analysis included them as naturally death.

There were similar situation happen in female rats too. The sponsor's analysis showed 38(60%), 39(60%), 35(54%), 26(41%), and 29(46%) survivors, while this reviewer's analysis showed 38

(58%), 39 (60%), 36 (55%), 27 (42%), and 30 (46%) in C1, C2, LD, MD, and HD group for female rats, respectively. These differences are due to the facts that there were seven female rats (#6846, #7025, #6742, and #6873 in LD; #6731, #6940, and #6780 in MD) that died naturally during their respective terminal sacrifice week (104⁺ week). These seven female rats lived up to their respective terminal sacrifice week and the 69% of tumors found in these female rats were benign. The tumors found in these female rats (#6742, #6873, #7025, #6780, and #6940) were caused the death. This reviewer classified these animals which the tumors did not cause the death as survivors, while the sponsor counted them as dead. Also, there were six female rats (#6969 and #6874 in C1, #6798 and #6766 in MD, and #6733 and #7072 in HD) that died accidently before week 94. The sponsor's analysis excluded these two animals while this reviewer's analysis included them as naturally death.

Based on the sponsor's report, the decreasing trend reached the statistical significance when compared with the pooled controls (p=0.0298) and C2 group individually (p=0.0414) in female rats. Due to the above survivors counting differences between the sponsor and this reviewer, based on the reviewer's analysis, the decreasing trend in female rats didn't reach the statistical significant when compared with the pooled controls (p=0.0594) and C2 group individually (p=0.0660). The pairwise comparisons show statistically significance increased mortality in the MD group compared to the pooled controls (p=0.0487) and compared to C2 (p=0.0363) in female rats.

3.2.2   Tumor Data Analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups. Both the dose response relationship tests and pairwise comparisons were performed using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993). In this method an animal that lives the full study period (w_{max}) or dies before the terminal sacrifice but develops the tumor type being tested gets a score of s_h =1. An animal that dies at week w_h without developing the tumor before the end of the study gets a score of $s_h = \left(\dfrac{w_h}{w_{max}} \right)^k$ <1. The adjusted group size is defined as Σs_h . As an interpretation, an animal with score s_h =1 can be considered as a whole animal while an animal with score s_h < 1 can be considered as a partial animal. The adjusted group size Σs_h is equal to N (the original group size) if all animals live up to the end of the study or if each animal that dies before the terminal sacrifice develops at least one tumor, otherwise the adjusted group size is less than N. These adjusted group sizes are then used for the dose response relationship (or the pairwise) tests using the Cochran-Armitage test. One critical point for Poly-k test is the choice of the appropriate value of k, which depends on the tumor incidence pattern with the increased dose. For long term 104 week standard rat and mouse studies, a value of k=3 is suggested in the literature. Hence, this reviewer used k=3 for the analysis of this data. For the calculation of p-values the exact permutation method was used. Note that, this review used the total number of animals in the groups as the denominator, assuming the animals that were not examined did not develop tumors.

The tumor rates and the p-values of the tested tumor types are listed in Tables 5A and 5B in the appendix for male and female rats, respectively.

Multiple testing adjustment: For the adjustment of multiple testing of dose response relationship, the FDA guidance for the carcinogenicity study design and data analysis suggests the use of test levels α=0.005 for common tumors and α=0.025 for rare tumors for a submission with two species, and a significance level α=0.01 for common tumors and α=0.05 for rare tumors for a submission with one specie in order to keep the false-positive rate at the nominal level of approximately 10%. A

File Name: outputfile792527110.pdf

rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. For multiple pairwise comparisons of treated group with control the FDA guidance the suggested the use of test levels $\alpha=0.01$ for common tumors and $\alpha=0.05$ for rare tumors, in order to keep the false-positive rate at the nominal level of approximately 10% for both submissions with two or one species.

It should be noted that the FDA guidance for multiple testing for dose response relationship is based on a publication by Lin and Rahman (1998). In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Lin and Rahman (2008) showed that this rule for multiple testing for dose response relationship is also suitable for Poly-K tests.

In their analysis the sponsor combined the tumor types as in the following table. This reviewer did the same and discussed with Dr. Behrsing, she had no additional tumor combinations.

Sex	Organ	Tumor		Tumor Presented as:
M	SYSTEMIC TUMORS	#B HEMANGIOMA #M HEMANGIOSARCOMA		HEMANGIOSARCOMA/ HEMANGIOMA
M/F	ADRENAL MEDULLA	#B PHEOCHROMOCYTOMA, BENIGN #M PHEOCHROMOCYTOMA, MALIGNANT		PHEOCHROMOCYTOMA MALIGNANT/BENIGN
M/F	LIVER	#B ADENOMA, HEPATOCELLULAR #M CARCINOMA, HEPATOCELLULAR		CARCINOMA/ADENOMA, HEPATOCELLULAR
M	KIDNEYS	#B ADENOMA #M CARCINOMA, RENAL TUBULE		CARCINOMA, RENAL TUBULE/ADENOMA
M/F	PANCREAS	#B ADENOMA, ISLET CELL #M CARCINOMA, ISLET CELL		CARCINOMA/ADENOMA ISLET CELL
M/F	PITUITARY	#B ADENOMA, PARS DISTALIS #M CARCINOMA, PARS DISTALIS		CARCINOMA/ADENOMA PARS DISTALIS
M/F	THYROID GLANDS	#B ADENOMA, C-CELL #M CARCINOMA, C-CELL		CARCINOMA/ADENOMA C-CELL
F	ADRENAL CORTEX	#B ADENOMA #M CARCINOMA		CARCINOMA/ADENOMA
F	BRAIN	#M ASTROCYTOMA, MALIGNANT, HIGH GRADE" #M ASTROCYTOMA, MALIGNANT, LOW GRADE		ASTROCYTOMA, MALIGNANT HIGH GRADE/LOW GRADE
F	MAMMARY GLAND	#B ADENOMA #M ADENOCARCINOMA #B FIBROADENOMA		ADENOCARCINOMA/ADENO MA/FIBROADENOMA
F	OVARIES	#B GRANULOSA CELL TUMOR, BENIGN #M THECOMA, MALIGNANT		THECOMA, MALIGNANT/GRANULOSA CELL TUMOR, BENIGN
F	SKIN	#B FIBROMA #M FIBROSARCOMA		FIBROSARCOMA/FIBROMA
F	STOMACH, NON	#B PAPILLOMA #B PAPILLOMA, SQUAMOUS		PAPILLOMA/PAPILLOMA, SQUAMOUS
F	THYROID GLANDS	#B ADENOMA, FOLLICULAR CELL #M CARCINOMA, FOLLICULAR CELL		CARCINOMA/ADENOMA FOLLICULAR CELL
F	UTERUS	#M CARCINOMA #M ADENOCARCINOMA		CARCINOMA/ADENOCARCI NOMA
M/F	SKIN SUBCUTIS	SKIN/SUBCUTIS	#M FIBROSARCOMA	#M FIBROSARCOMA
M/F	SKIN SUBCUTIS	SKIN/SUBCUTIS	#B FIBROMA #M FIBROSARCOMA	FIBROSARCOMA/FIBROMA

#B = benign tumor; #M = malignant tumor

Reviewer's findings: Following two tables display the tumor types showed p-values less than or equal to 0.05 either for dose response relationship or pairwise comparisons of treated groups and control.

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups and Combined Controls in Male Rats

File Name: outputfile792527110.pdf

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value			
						Dose Response	C vs. LD	C vs. MD	C vs. HD
ADRENAL MEDULLA	#M PHEOCHROMOCYTOMA,	1	4	1	0	0.8376	0.0419*	0.5270	0 3279
	C_PHEOCHROMOCYTOMA_M+B	13	15	7	10	0.3156	0.0085*	0.4106	0.1677
LIVER	#B ADENOMA, HEPATOCE	2	2	3	6	0.0080	0.3976	0.1733	0 0164
	C_HEPATOCELLULAR_M+B	3	2	3	6	0.0153	0.5254	0.2675	0 0351
PAWS	#B PAP LLOMA	0	0	0	3	0.0081*	.	.	0.0352*
SYSTEMIC TUMORS	#B HEMANGIOMA	1	0	2	3	0.0267	0.3279	0.2278	0.1026
	C_HEMANGIO_M+B	1	0	2	4	0.0077*	0.3279	0.2278	0.0393*
Multi/Organ**	C_PAPILLOMA_B	4	0	4	3	0.1675	0.7976	0.2026	0.4230

For dose response, *Indicted the significant at 0.005 or 0 025 alpha levels; for pairwise comparison, *Indicted the significant at 0.01 or 0 05 alpha levels.
**: Combined all Papilloma tumors in organ of Ear, tail, skin, and paw.

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups and Combined Controls in Female Rats

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value			
						Dose Response	C vs. LD	C vs. MD	C vs. HD
CERVIX	#B GRANULAR CELL TUM	0	2	2	5	0.0029*	0.1139	0.1083	0 0030*
VAGINA	#B GRANULAR CELL TUM	0	0	0	3	0.0064*	.	.	0 0309*
CERVIX/ VAGINA	C_GRANULAR_CELL_B	0	2	2	8	<0 001*	0.1139	0.1083	<0.001*
EARS	#B NEURAL CREST TUMO	0	0	0	2	0 0352	.	.	0.0998
K DNEYS	#B LIPOMA	0	0	1	2	0 0274	.	0.3269	0.0998
LIVER	C_HEPATOCELLULAR_M+B	2	2	3	4	0 0461	0.4187	0.2035	0.0817
OVAR ES	#B GRANULOSA CELL TU	0	1	1	3	0.0129*	0 3396	0.3269	0 0309*
	C_GRANULOSA_THECOMA	1	1	1	3	0 0401	0 5653	0.5484	0.0953
PITUITARY	C_PARS_DISTALIS_M+B	97	54	48	56	0 0608	0.1429	0.4794	0.0466
SK N	#M FIBROSARCOMA	1	1	0	4	0 0105	0 5653	0.3269	0.0377
	C_F BROMA_M+B	1	2	0	4	0 0200	0 2663	0.3269	0.0377
SK N/SUBCUTIS	C_F BROSARCOMA_M	1	2	1	4	0 0217	0 2719	0.5484	0.0377
	C_F BROSARCOMA_M+B	1	3	2	4	0 0396	0.1177	0.2550	0.0377
THYRO D GLANDS	#M CARCINOMA, C-CELL	1	4	1	0	0 8309	0 0458	0.5484	0.3182
UTERUS	#B POLYP	4	10	5	5	0 2678	0.0036*	0.1349	0.1158

For dose response, *Indicted the significant at 0 005 or 0.025 alpha levels; for pairwise comparison, *Indicted the significant at 0 01 or 0.05 alpha levels.

In male rats, tests on tumor data showed a statistically significant positive trend (p=0.0077 <0.025) in incidence of benign+maglinant hemangioma as a systemic tumor; the pairwise comparison showed statistically significant increased incidence in HD group in this tumor type compared to the combined controls (p=0.0393 <0.05). Tests on tumor data also showed a statistically significant positive trend (p = 0.0081 <0.025) in incidence of benign papilloma tumor in paws; the pairwise comparison showed statistically significant increased incidence in HD group in this tumor type compared to the combined controls (p=0.0352 <0.05). But, the combination of all Papilloma tumors in organ of Ear, tail, skin, and paw did not show any statistically significant in trend test and pairwise comparison. The pairwise comparison showed statistically significant increased incidence of malignant pheochromocytoma in adrenal in LD group compared to the combined controls (p=0.0419<0.05) and the combination (benign+malignant) of pheochromocytoma malignant and benign in LD group compared to the combined controls (p=0.0085<0.01). However, the dose response relationships for these tumor types were not statistically significant.

File Name: outputfile792527110.pdf

In female rats, test on tumor data showed a statistically significant positive trend (p<0.025) in following tumor types: benign granular cell tumors in cervix (p=0.0029), in vagina (p=0.0064), and combination of two organs (p<0.001); benign granulosa cell tumors in ovaries (p=0.0129); and malignant fibrosarcoma (p=0.0105), benign+malignant fibroma (p=0.0200) in skin, and malignant fibrosarcoma in skin/subcutis (p=0.0217). The pairwise comparison also showed statistically significant increased incidence in HD group compared to the combined controls (with p values less than 0.05). These tumors included benign granular cell tumor in cervix (p=0.0030) and in vagina (p=0.0309), and combination of two organs (p<0.001); benign granulosa cell tumor in ovaries (p=0.0309), malignant fibrosarcoma (p=0.0377) and fibrosarcoma tumor (benign or malignant) in skin (p=0.0377), and malignant fibrosarcoma (p=0.0377) and malignant fibrosarcoma (p=0.0377) in skin/subcutis. The pairwise comparison also showed statistically significant increased incidence of malignant carcinoma c-cell in thyroid glands in LD group compared to the combined controls (p=0.0458<0.05) and the benign polyp in uterus in LD group compared to the combined controls (p=0.0036<0.01). However, the dose response relationships for these tumor types were not statistically significant.

4   Mouse Study

Study Report: (b) (4) **661038.pdf; SAS data: 661038ft.xpt and 661038mt.xpt**

This study assessed the carcinogenic potential of INT-747 in male and female CD-1 mice. The test material was administered at doses of 4, 10 or 25 mkd of INT-747 once daily by oral gavage for at least 104 weeks. This review refers these dose groups as the low (LD), mid (MD), and high (HD) dose groups, respectively. There were two control groups; the control animals received the vehicle (0.5% carboxymethylcellulose [CMC] in deionized water). The dose volume for all groups was 10 mL/kg. There were 65 mice/sex/dose. In addition, 9 mice/sex in one control group and 40 mice/sex in the treated groups served as toxicokinetic animals.

Activities performed included clinical examinations (twice daily examinations for mortality and signs of ill health or reaction to dosing, daily post dosing cage side observations, weekly detailed examinations, palpable mass examinations on main study animals), body weights, food consumption, laboratory investigations (hematology) for health screen animals prior to dosing initiation and for main animals at study termination (blood smears). Post-mortem evaluations for main study animals included macroscopic observations and microscopic observations.

4.1   Sponsor's Analyses

4.1.1   Survival Analysis

The sponsor performed survival analysis using the same methodologies that were used in the rat study.

Sponsor's findings: The sponsor reported the total unscheduled deaths in the following table. Male and female control groups had similar survival rates; the survival rates ranged from 32% to 45% for males and 42% to 63% in females. There was an INT-747-related higher survival rate in the high-dose (25 mkd) group females compared to controls (p=0.0098). Statistical evaluation showed that the difference in the females was statistically significant in the trend test (p=0.0095) when using the pooled controls and C1 (p=0.0240) and 2 individually (p=0.0036). The sponsor stated that this higher survival rate in the 25 mg/kg/day group females was considered INT-747-related and most likely related to the more than 20% lower cumulative body weight gain from interval 0 (Leakey, 2004). There were no statistical differences in survival among the male groups of animals. Among the males

File Name: outputfile792527110.pdf

and females, at least 33 mice per group survived into Study Week 90 suggesting that exposure to INT-747 was adequate for informative interpretation of the tumor incidence results.

The Sponsor's Report of Total Unscheduled Deaths, n=65/sex/dose group)

Dosage (mg/kg/day)	0	0	4	10	25
Males [a]	65	65	65	65	65
Number survival	21	29	25	21	25
Number of animals found dead or euthanized in extremis	44	35	39	43	40
Accidental deaths	0	1	1	1	0
% survival [b]	32%	45%	39%	33%	38%
Females [a]	65	65	65	65	65
Number survival	27	24	25	28	41
Number of animals found dead or euthanized in extremis	38	40	40	36	24
Accidental deaths	0	1	0	1	0
% survival [b]	42%	38%	38%	44%	63%

[a] = number of animals per dose group
[b] = % survival was corrected based on the number of accidental deaths

[Source: page 46 of study report of (b) (4) 661038.pdf]

4.1.2   Tumor Data Analysis

The sponsor analyzed the tumor incidence data using the Peto's mortality-prevalence method, without continuity correction, incorporating the context (incidental or fatal) in which tumors were observed. The following fixed intervals were used for incidental tumor analyses: weeks 0-50, 51-80, 81-end of study, and scheduled terminal sacrifice. For each sex, tumors that were detected, either by palpation or necropsy, after the first animal of that sex was terminally sacrificed were considered incidental and included in the scheduled terminal sacrifice interval for analyses. For example, tumors that were detected in male mice after the first sacrifice on study day 728 were considered incidental and included in the scheduled terminal sacrifice interval for the purpose of statistical analysis. Tumors classified as mortality-independent, such as, but not limited to, those of the mammary gland and skin, were analyzed with Peto's mortality-independent method incorporating the day of detection.

Adjustment for multiple testing: The sponsor used the same methodologies for the adjustment for multiple testing.

Sponsor's findings:
There was a statistically significant positive trend in the incidence of hepatocellular carcinoma in the females when compared to the C1 group. However, statistical significance was limited to the trend test and compared to only one of the control groups and no statistical significance (pairwise comparison or trend tests) was seen with hepatocellular adenoma or when all hepatocellular adenomas and carcinomas were combined. The sponsor considered this finding as incidental and unrelated to INT-747 administration.

4.2   Reviewer's Analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing pharmacologist, this reviewer performed survival and tumor data analyses using data submitted electronically in NDA 207-999 on 12/19/2014.

File Name: outputfile792527110.pdf

4.2.1 Survival Analysis

The Kaplan-Meier curves for survival rates of all treatment groups are given in Figures 2A and 2B in the appendix for male and female mice, respectively. The intercurrent mortality data of all treatment groups are given in Tables 2A and 2B in the appendix for male and female mice, respectively. Results of the tests for dose response relationship and homogeneity of survivals for control, low, medium, and high dose groups are given in Tables 4A and 4B in the appendix for male and female mice, respectively.

Reviewer's findings: This reviewer's analysis showed the numbers (percent) of survival were 22(34%), 30(46%), 26(40%), 22(34%), and 25(38%) in male mice and 27(42%), 24(37%), 25(38%), 28 (43%), and 41(63%) in female mice in the C1, C2, LD, MD, and HD groups, respectively. There was a statistically significant increase in the survival rate of the high-dose (25 mkd) group females when pairwise comparisons were made with both the pooled female control groups and C2 group individually. Statistical evaluation showed that the difference in the females was statistically significant in the trend test (p=0.0021) when using the pooled controls and C1 (p=0.0064) and 2 individually (p=0.0009).

Reviewer's comment: The sponsor's analysis showed 21(32%), 29(45%), 25(39%), 21(33%), and 25(38%) survivors, while this reviewer's analysis showed 22(34%), 30(46%), 26(40%), 22(34%), and 25(38%) in C1, C2, LD, MD, and HD group for male mice, respectively. These differences are due to the facts that there were seven male mice (#8427 in C1, #8200 and #8397 in C2, #8026 in LD, #8128 and #8315 in MD, and #8033 in HD) that died naturally during their respective terminal sacrifice week(104+ week). These seven male mice lived up to their respective terminal sacrifice week (104+ week) and the tumors found in these male mice were only 15% labeled as malignant. The tumors found in these male mice (#8200, #8033, and #8315) were labeled as malignant and caused the death. This reviewer classified the animals which the tumors did not cause the death as survivors, while the sponsor counted them as dead. There were two male mice (#8156 in C2, #8013 in LD and #8036 in MD) that died accidently before week 28. The sponsor's analysis excluded these two animals while this reviewer's analysis included them as naturally death.

There were similar situation happen in female mice too. The sponsor's analysis showed 27(42%), 24(38%), 25(38%), 28(44%), and 41(63%) survivors, while this reviewer's analysis showed 27(42%), 24(37%), 25(38%), 29 (45%), and 41(63%) in C1, C2, LD, MD, and HD group for female mice, respectively. These differences are due to the facts that there was one female mouse (#8890 in MD) that died naturally during their respective terminal sacrifice week (104 week). The tumors found in this female mouse was labeled as malignant and caused the death. This reviewer classified this animal as dead, while the sponsor counted them as dead (agreement between the sponsor and reviewer). This reviewer classified these animals as death at week 103, survivors, while the sponsor counted them as dead (disagreement between the sponsor and reviewer). Also, there were two female mice (#8944 in C2, #8927 in MD) that died accidently before week 80. The sponsor's analysis excluded these two animals while this reviewer's analysis included them as naturally death.

Both the sponsor and this reviewer reached the same conclusion for the survival analysis.

4.2.2 Tumor Data Analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups using the same method that was used for the rat study. The tumor rates and the p-values of the tested tumor types are listed in Tables 6A and 6B in the appendix for male and female mice, respectively.

File Name: outputfile792527110.pdf

The sponsor combined the tumor incidences as follow. This reviewer did the same and discussed with Dr. Behrsing, she had no additional tumor combinations.

Sex	Organ Name	Tumor Name	Tumor Name	Combined Tumor Name
F	Harderian Glands	#B ADENOMA	#B ADENOMA, MULT PLE	#B ADENOMA
M/F	Liver	#B ADENOMA, HEPATOCELLULAR	#B ADENOMA, HEPATOCELLULAR, MULTIPLE	#B ADENOMA, HEPATOCELLULAR
		#M ADENOMA, HEPATOCELLULAR	#M ADENOMA, HEPATOCELLULAR, MULTIPLE	#M ADENOMA, HEPATOCELLULAR
		#B ADENOMA, HEPATOCELLULAR	#M ADENOMA, HEPATOCELLULAR	CARCINOMA/ADENOMA, HEPATOCELLULAR
M/F	Lung	#B ADENOMA, BRONCHIOLO-ALVEOLAR	#B ADENOMA, BRONCHIOLO-ALVEOLAR, MULTIPLE	#B ADENOMA, BRONCHIOLO-ALVEOLAR
		#B ADENOMA, BRONCHIOLO-ALVEOLAR	#B ADENOMA, BRONCHIOLO-ALVEOLAR, MULTIPLE	#B ADENOMA, BRONCHIOLO-ALVEOLAR
		#B ADENOMA, BRONCHIOLO-ALVEOLAR	#M ADENOMA, BRONCHIOLO-ALVEOLAR	CARCINOMA/ADENOMA, BRONCHIOLO-ALVEOLAR
M	Testes	#B ADENOMA, LEYDIG CELL	#B ADENOMA, LEYDIG CELL, MULTIPLE	#B ADENOMA, LEYDIG CELL
F	Harderian Gland	#B ADENOMA	#M CARC NOMA	CARCINOMA/ADENOMA
	Systemic Tumors	#B HEMANGIOMA	#M HEMANGIOSARCOMA	HEMANGIOSARCOMA/ HEMANGIOMA
	Lung	#B ADENOMA, BRONCHIOLO-ALVEOLAR	#B ADENOMA, BRONCHIOLO-ALVEOLAR, MULTIPLE	#B ADENOMA, BRONCHIOLO-ALVEOLAR
		#M ADENOMA, BRONCHIOLO-ALVEOLAR	#M ADENOMA, BRONCHIOLO-ALVEOLAR, MULTIPLE	#M ADENOMA, BRONCHIOLO-ALVEOLAR

Sex	Organ Name	Organ Name	Tumor Name	Combined Organ Name
M/F	SKIN	SUBCUTIS	#M F BROSARCOMA	SK N/SUBCUTIS
F	CERVIX	UTERUS	#B POLYP, ENDOMETRIAL STROMAL	CERVIX/UTERUS

Reviewer's findings: Following two tables display the tumor types showed p-values less than or equal to 0.05 either for dose response relationship or pairwise comparisons of treated groups and control.

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups and Combined Controls in Male Mice

Organ Name	Tumor Name	0 mg/kg/day C N=130	4 mg/kg/day LD N=65	10 mg/kg/day MD N=65	25 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
SUBCUTIS	#M F BROSARCOMA	1	0	0	3	0.0312	0.3309	0.3309	0.1232

For dose response, *Indicted the significant at 0 005 or 0.025 alpha levels; for pairwise comparison, *Indicted the significant at 0 01 or 0.05 alpha levels.

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups and Combined Controls in Female Mice

Organ Name	Tumor Name	0 mg/kg/day C N=130	4 mg/kg/day LD N=65	10 mg/kg/day MD N=65	25 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
CERVIX	#B POLYP, ENDOMETRIA	0	0	0	2	0.0493	.	.	0.1302
HARDERIAN GLAND	#B ADENOMA	2	0	3	4	0.0341	0 5408	0 2170	0.1272
LIVER	#M CARC NOMA, HEPATO	1	0	1	4	0 0119*	0 3212	0 5603	0 0584
SYSTEMIC TUMORS	#M HEMANGIOSARCOMA	8	9	4	3	0.8472	0 0493	0 3718	0 6023

For dose response, *Indicted the significant at 0 005 or 0.025 alpha levels; for pairwise comparison, *Indicted the significant at 0 01 or 0.05 alpha levels.

File Name: outputfile792527110.pdf

In female mice, test on tumor data showed a statistically significant increasing trend (p=0.0119<0.025) in carcinoma, hepato in liver; the pairwise comparison show numerically increased incidence in HD group compared to the combined controls (p=0.0584); the p-value was slightly over 0.05 alpha levels.

5 Conclusion

This review evaluates statistically the tumorigenicity data of 2-year oral carcinogenicity studies of INT-747 in rats and mice. The review analyzes the dose-response relationship of tumor incidence and mortality (including tumor-related mortality). Tumor analyses consisted of trend analyses for dose-response relationship in tumor incidence and pairwise comparisons in tumor incidences of individual treatment groups with control. The review concludes that INT-747 decreased survival in female rats and increased tumor incidences in several organs in male and female rats. The male rats showed increases in the incidence of total hemangioma and papilloma in the paws; but the combination of all papilloma tumors in organ of ear, tail, skin, and paw did not show any statistically significant in trend analyses for dose-response and pairwise comparison between any treatment groups with controls. The female rats showed increases in the incidence of granular cell tumors in the reproductive system (cervix and vagina), benign granulosa cell tumor in ovaries, and fibrosarcoma in the skin. The drug had no effects on survival or tumor incidence in either sex in mice.

Rat Study: Rats (65/sex/dose) were dosed by oral gavage with INT-747 daily for up to 104 weeks. The INT-747 dose was 0, 0, 2, 7, or 20-mg/kg in the control 1 (C1), control 2 (C2), low-dose (LD), mid-dose (MD), and high-dose (HD) groups, respectively, in both males and females. The control groups (0 mg/kg/day) received the vehicle only.

Survival analyses showed dose-related significant decreases in survival rates in females, but not in males. The respective survival rate in C1, C2, LD, MD, and HD groups at terminal sacrifice were 58%, 60%, 55%, 42%, and 46% in females (p=0.0594) and 31%, 28%, 28%, 26%, and 25% in males. The pairwise comparisons show statistically significant decreases in survival rates in MD group compared to the pooled controls (p=0.0487) and compared to C2 group (p=0.0363) in female rats.

Tumor analyses showed positive responses in both trend analyses and pairwise comparisons in tumor incidences in both sexes. In males, the trend analysis showed statistically significant increase in the incidence of total hemangioma (both benign and malignant combined) (p= 0.0077) and benign papilloma tumor in paws (p=0.0081). The pairwise comparison showed that only the HD group had statistically significant increases in tumor incidence in these two tumor types when compared to the pooled controls (p= 0.0393, and p= 0.0352, respectively). The combination of all papilloma tumors in organ of ear, tail, skin, and paw did not show any statistically significant in analyses for dose-response and pairwise comparison between any treatment groups with controls.

In female rats, tests on tumor data showed a statistically significant positive trend (with p<0.025) in following tumor types: benign granular cell tumors in cervix (p=0.0029), in vagina (p=0.0064), and combination of the two organs (p<0.001); benign granulosa cell tumors in ovaries (p=0.0129); and malignant fibrosarcoma (p=0.0105), benign+malignant fibroma (p=0.0200) in skin, and malignant fibrosarcoma in skin/subcutis (p=0.0216). The pairwise comparisons also showed statistically significant increased incidence in HD group compared to the combined controls (with p values less than 0.05). These tumors included benign granular cell tumor in cervix (p=0.0030) and in vagina

File Name: outputfile792527110.pdf

(p=0.0309), and combination of two organs (p<0.001); benign granulosa cell tumor in ovaries (p=0.0309), malignant fibrosarcoma (p=0.0377) and fibrosarcoma tumor (benign or malignant) in skin (p=0.0377), and malignant fibrosarcoma (p=0.0377) and malignant fibrosarcoma (p=0.0377) in skin/subcutis.

Mouse Study: Mice (65/sex/dose) were dosed by oral gavage with INT-747 daily for up to 104 weeks. The INT-747 dose was 0, 0, 4, 10, or 25-mkd in the C1, C2, LD, MD, and HD groups, respectively, in both sexes.   The control groups (0 mg/kg/day) received the vehicle only.

The survival analysis showed that the HD females had a statistically significant decrease in mortality when compared to the pooled controls. The trend test and pairwise comparisons did not show statistically significant increased mortality in any treated group in male mice. The respective survival rates in the C1, C2, LD, MD and HD groups at the termination (Week 104) were 42%, 37%, 38%, 43%, and 63% in females and 34%, 46%, 40%, 34%, and 38% in males .

In female mice, tests on tumor data showed a statistically significant increasing trend (p=0.0119<0.025) in carcinoma, hepato in liver; the pairwise comparison show a numerically increased incidence in HD group compared to the combined controls (p=0.0584); the p-value was slightly over 0.05 alpha levels.

<div align="right">

Feng Zhou, M.S.
Mathematical Statistician

</div>

Secondary Reviewer: Atiar Mohammad Rahman, Ph.D.
Secondary Reviewer: Karl Lin, Ph.D., Team Leader, Biometrics-VI

cc:
Dr. Tracy Behrsing
Dr. Sushanta Chakder
Dr. Tsong
Ms. Zhou
Dr. Lin
Ms. Patrician

File Name: outputfile792527110.pdf

6 Appendix

Table 1A: Intercurrent Mortality – Male Rats

Week	0 mg/kg/day (n=65)		0 mg/kg/day (n=65)		2 mg/kg/day (n=65)		7 mg/kg/day (n=65)		20 mg/kg/day (n=65)	
	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	2	3 08	8	12 31	9	13.85	6	9 23	4	6.15
53 - 78	19	32.31	19	41 54	14	35.38	19	38.46	22	40.00
79 - 91	12	50.77	15	64 62	15	58.46	19	67.69	14	61.54
92 - 100	12	69.23	5	72 31	9	72.31	4	73.85	9	75.38
Ter. Sac.	20	30.77	18	27 69	18	27.69	17	26.15	16	24.62

Cum. %: Cumulative percentage except for Ter. Sac.

Table 1B: Intercurrent Mortality - Female Rats

Week	0 mg/kg/day (n=65)		0 mg/kg/day (n=65)		2 mg/kg/day (n=65)		7 mg/kg/day (n=65)		20 mg/kg/day (n=65)	
	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	2	3 08	1	1.54	.	.	2	3 08	1	1 54
53 - 78	12	21.54	6	10.77	9	13.85	5	10.77	11	18.46
79 - 91	9	35.38	7	21 54	7	24.62	11	27.69	13	38.46
92 – 103	4	41.54	12	40 00	13	44.62	20	58.46	10	53.85
Ter. Sac.	38	58.46	39	60 00	36	55.38	27	41.54	30	46.15

Cum. %: Cumulative percentage except for Ter. Sac.

Table 2A: Intercurrent Mortality – Male Mice

Week	0 mg/kg/day (n=65)		0 mg/kg/day (n=65)		4 mg/kg/day (n=65)		10 mg/kg/day (n=65)		25 mg/kg/day (n=65)	
	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	3	4 62	3	4.62	6	9.23	3	4 62	1	1 54
53 - 78	17	30.77	13	24 62	14	30.77	14	26.15	12	20.00
79 - 91	15	53.85	7	35 38	8	43.08	11	43.08	12	38.46
92 - 103	8	66.15	12	53 85	11	60.00	15	66.15	15	61.54
Ter. Sac.	22	33.85	30	46.15	26	40.00	22	33.85	25	38.46

Cum. %: Cumulative percentage except for Ter. Sac.

Table 2B: Intercurrent Mortality – Female Mice

Week	0 mg/kg/day (n=65)		0 mg/kg/day (n=65)		4 mg/kg/day (n=65)		10 mg/kg/day (n=65)		25 mg/kg/day (n=65)	
	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	3	4 62	6	9.23	7	10.77	2	3 08	3	4 62
53 - 78	9	18.64	12	27 69	8	23.08	11	20.00	6	13.85
79 - 91	7	29.23	9	41 54	17	49.23	13	40.00	8	26.15
92 - 103	19	58.46	14	63 08	8	61.54	11	56.92	7	36.92
Ter. Sac.	27	41.54	24	36 92	25	38.46	28	43.08	41	63.08

Cum. %: Cumulative percentage except for Ter. Sac.

Table 3A: Intercurrent Mortality Comparison – Male Rats

Test	Statistic	P-Value (Pooled Controls)	P-Value (Control-1)	P-Value (Control-2)
Dose-Response	Likelihood Ratio	0.4597	0 3545	0 8408

File Name: outputfile792527110.pdf

Test	Statistic	P-Value (Pooled Controls)	P-Value (Control-1)	P-Value (Control-2)
Homogeneity	Log-Rank	0.7587	0 5331	0 9596

Table 3B: Intercurrent Mortality Comparison – Female Rats

Test	Statistic	P-Value (Pooled Controls)	P-Value (Control-1)	P-Value (Control-2)
Dose-Response	Likelihood Ratio	0 0594	0.1706	0 0660
Homogeneity	Log-Rank	0.1310	0 3554	0.1250

Table 4A: Intercurrent Mortality Comparison – Male Mice

Test	Statistic	P-Value (Pooled Controls)	P-Value (Control-1)	P-Value (Control-2)
Dose-Response	Likelihood Ratio	0 8332	0 3703	0.7072
Homogeneity	Log-Rank	0 9026	0 6596	0 6026

Table 4B: Intercurrent Mortality Comparison – Female Mice

Test	Statistic	P-Value (Pooled Controls)	P-Value (Control-1)	P-Value (Control-2)
Dose-Response	Likelihood Ratio	0 0021	0 0064	0 0009
Homogeneity	Log-Rank	0 0174	0 0264	0 0136

Table 5A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Male Rats

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
ADIPOSE TISSUE	#B HIBERNOMA, BENIGN	1	0	0	0	0.5850	0.3252	0.3083	0.3252
	#B LIPOMA	1	0	1	1	0.2614	0.3252	0.5234	0.5464
ADRENAL CORTEX	#B ADENOMA	4	0	1	2	0.3746	0.7976	0.4909	0.6489
ADRENAL MEDULLA	#B PHEOCHROMOCYTOMA,	12	11	6	10	0.1724	0.0751	0.4905	0.1274
	#M PHEOCHROMOCYTOMA,	1	4	1	0	0.8376	0.0419*	0.5270	0.3279
	C_PHEOCHROMOCYTOMA_M+B	13	15	7	10	0.3156	0.0085*	0.4106	0.1677
BRAIN	#M ASTROCYTOMA, MALI	1	0	0	1	0.3672	0.3252	0.3083	0.5538
CAVITY, THORACI	#M HIBERNOMA, MALIGN	0	1	0	0	0.3850	0.3333	.	.
DUODENUM	#M ADENOCARCINOMA	0	0	1	0	0.3869	.	0.3109	.
EARS	#B PAPILLOMA	0	0	1	0	0.3869	.	0.3109	.
HEART	#B SCHWANNOMA, ENDOC	1	0	1	0	0.4713	0.3279	0.5270	0.3279
K DNEYS	#B ADENOMA	1	1	0	0	0.6649	0.5501	0.3109	0.3279
	#M CARCINOMA, RENAL	0	0	0	1	0.2010	.	.	0.3279
	C_RENAL-TUBULE_M+B	1	1	0	1	0.3892	0.5501	0.3109	0.5501
LIVER	#B ADENOMA, HEPATOCE	2	2	3	6	0.0080	0.3976	0.1733	0.0164
	#B CHOLANGIOMA	1	0	0	0	0.5879	0.3279	0.3109	0.3279
	#M CARCINOMA, HEPATO	1	0	0	0	0.5850	0.3252	0.3083	0.3252
	C_HEPATOCELLULAR_M+B	3	2	3	6	0.0153	0.5254	0.2675	0.0351
LUNGS	#B ADENOMA, BRONCHIO	1	0	0	1	0.3608	0.3252	0.3083	0.5464
MAMMARY GLAND	#B ADENOMA	1	1	0	0	0.6649	0.5501	0.3109	0.3279
PANCREAS	#B ADENOMA, ACINAR C	1	1	0	1	0.3892	0.5501	0.3109	0.5501

File Name: outputfile792527110.pdf

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value			
						Dose Response	C vs. LD	C vs. MD	C vs. HD
	#B ADENOMA, ISLET CE	10	8	5	6	0.4413	0.1918	0.5086	0.4335
	#B LIPOMA	1	0	0	0	0.5879	0.3279	0.3109	0.3279
	#M ADENOCARCINOMA, D	0	0	1	0	0.3869	.	0.3109	.
	#M CARCINOMA, ISLET	1	1	1	0	0.5830	0.5501	0.5270	0.3279
	#M SARCOMA, UND FFER	1	0	0	0	0.5850	0.3252	0.3083	0.3252
	C_ISLET-CELL_M+B	11	9	6	6	0.5172	0.1566	0.4439	0.5072
PARATHYROIDS	#B ADENOMA	3	1	1	2	0.3470	0.3976	0.3671	0.5418
PAWS	#B OSTEOCHONDROMA	1	0	0	0	0.5879	0.3279	0.3109	0.3279
	#B PAPILLOMA	0	0	0	3	0.0081*	.	.	0.0352*
PITUITARY	#B ADENOMA, PARS DIS	74	34	32	37	0.3560	0.4501	0.5181	0.4378
	#M CARCINOMA, PARS D	0	1	0	0	0.3850	0.3333	.	.
	C_PARS-DISTALIS_M+B	74	35	32	37	0.3810	0.4975	0.5181	0.4378
SAL. GLAND MAND	#M ADENOCARCINOMA	1	0	0	0	0.5850	0.3252	0.3083	0.3252
SEM NAL VESICLE	#M ADENOCARCINOMA	1	0	0	0	0.5879	0.3279	0.3109	0.3279
SKELETAL MUSCLE	#M F BROSARCOMA	0	0	1	0	0.3900	.	0.3167	.
SK N	#B KERATOACANTHOMA,	3	2	2	2	0.3707	0.5364	0.4909	0.5254
	#B LIPOMA	1	0	0	0	0.5850	0.3252	0.3083	0.3252
	#B PAPILLOMA	3	0	3	0	0.7588	0.7000	0.2725	0.7000
	#M CARCINOMA	0	0	1	0	0.3869	.	0.3109	.
	#M CARCINOMA, SQUAMO	1	0	0	0	0.5879	0.3279	0.3109	0.3279
	#M F BROSARCOMA	0	2	0	1	0.2907	0.1129	.	0.3333
	#M OSTEOSARCOMA	1	0	0	0	0.5879	0.3279	0.3109	0.3279
SK N/SUBCUTIS	C_FIBROSARCOMA_M	2	3	0	2	0.3809	0.2098	0.5234	0.4026
	C_FIBROSARCOMA_M+B	6	4	2	2	0.7217	0.4407	0.4792	0.5217
SOFT TISSUE-TH	#B HIBERNOMA, BENIGN	2	2	1	3	0.1331	0.4074	0.6766	0.2057
	#M F BROSARCOMA	0	0	1	0	0.3900	.	0.3167	.
	#M HIBERNOMA, MALIGN	0	2	1	0	0.5567	0.1129	0.3167	.
STOMACH, NON	#B PAPILLOMA	1	0	0	0	0.5850	0.3252	0.3083	0.3252
SUBCUTIS	#B F BROMA	4	1	2	0	0.8722	0.5254	0.6147	0.7976
	#B KERATOACANTHOMA,	1	0	0	0	0.5879	0.3279	0.3109	0.3279
	#B LIPOMA	0	1	0	1	0.1960	0.3333	.	0.3279
	#M F BROSARCOMA	2	1	0	1	0.4583	0.6964	0.5234	0.6964
	#M OSTEOSARCOMA	1	0	0	0	0.5879	0.3279	0.3109	0.3279
	#M SARCOMA, SP NDLE	0	1	1	0	0.3943	0.3333	0.3109	.
SYSTEMIC TUMORS	#B HEMANGIOMA	1	0	2	3	0.0267	0.3279	0.2278	0.1026
	#M HEMANGIOSARCOMA	0	0	0	1	0.2010	.	.	0.3279
	#M LYMPHOMA, MALIGNA	3	0	2	0	0.7580	0.6928	0.4974	0.6928
	#M MESOTHELIOMA, MAL	0	1	0	0	0.3850	0.3333	.	.
	#M SARCOMA, HISTIOCY	9	0	2	1	0.9108	0.9731	0.7212	0.8895
	C_HEMANGIO_M+B	1	0	2	4	0.0077*	0.3279	0.2278	0.0393*
TAIL	#B ADENOMA, SEBACEOU	0	0	1	0	0.3869	.	0.3109	.
	#B OSTEOMA	0	0	0	1	0.2010	.	.	0.3279
	#B PAPILLOMA, SQUAMO	2	0	0	0	0.8314	0.5501	0.5270	0.5501
	#M OSTEOSARCOMA	0	0	0	1	0.2010	.	.	0.3279
TESTES	#B ADENOMA, LEYDIG C	3	3	3	2	0.4271	0.3058	0.2837	0.5309
THYMUS	#M HIBERNOMA, MALIGN	0	0	0	1	0.2010	.	.	0.3279

File Name: outputfile792527110.pdf

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value			
						Dose Response	C vs. LD	C vs. MD	C vs. HD
THYRO D GLANDS	#B ADENOMA, C-CELL	15	9	6	6	0.7065	0.3757	0.5019	0.5485
	#B ADENOMA, FOLLICUL	8	4	3	2	0.8023	0.5911	0.4577	0.6779
	#M CARCINOMA, C-CELL	0	1	0	0	0.3869	0.3279	.	.
	C_C-CELL_M+B	15	9	6	6	0.7065	0.3757	0.5019	0.5485
ZYMBAL'S GLANDS	#B CYSTADENOMA	1	0	0	0	0.5850	0.3252	0.3083	0.3252
	#M ADENOCARCINOMA	1	0	0	0	0.5850	0.3252	0.3083	0.3252
Multi/Organ**	C_PAP LLOMA_B	4	0	4	3	0.1675	0.7976	0.2026	0.4230

**: Combined all Papilloma tumors in organ of Ear, tail, skin, and paw.

Table 5B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Female Rats

APPEARS THIS WAY ON ORIGINAL

	#B POLYP	0	0	1	0	0.3861	.	0.3269	.
	#M ADENOCARCINOMA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	#M LEIOMYOSARCOMA	1	0	1	0	0.4638	0 3396	0.5484	0.3182
	#M SARCOMA, ENDOMETR	4	2	0	0	0.9655	0 3301	0.7986	0.7879
	#M SCHWANNOMA, MALIG	1	1	0	0	0.6669	0 5653	0.3269	0.3182
CERVIX/ VAGINA	C_ GRANULAR CELL_B	0	2	2	8	<0.001*	0.1139	0.1083	<0.001*
DUODENUM	#M LEIOMYOSARCOMA	0	1	0	0	0.3861	0 3396	.	.
EARS	#B NEURAL CREST TUMO	0	0	0	2	0.0352	.	.	0.0998
G NGIVA	#B PAPILLOMA	0	0	1	0	0.3861	.	0.3269	.
ILEUM	#M ADENOCARCINOMA	0	1	0	0	0.3861	0 3396	.	.
JEJUNUM	#B LEIOMYOMA	0	1	0	1	0.1892	0 3396	.	0.3182
KIDNEYS	#B LIPOMA	0	0	1	2	0.0274	.	0.3269	0.0998
	#M NEPHROBLASTOMA	1	0	0	0	0.5923	0 3375	0.3248	0.3161
LIVER	#B ADENOMA, HEPATOCE	2	2	2	3	0.1145	0.4187	0.3959	0.1845
	#M CARCINOMA, HEPATO	0	0	1	1	0.1106	.	0.3312	0.3182
	C_HEPATOCELLULAR_M+B	2	2	3	4	0.0461	0.4187	0.2035	0.0817
LUNGS	#B EPITHELIOMA, CYST	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	#M ADENOCARCINOMA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
MAMMARY GLAND	#B ADENOMA	17	7	6	5	0.8280	0 6131	0.6987	0.7674

File Name: outputfile792527110.pdf

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
	#B F BROADENOMA	42	17	18	15	0.8366	0 8290	0.6506	0.8479
	#B F BROMA	0	1	0	0	0.3861	0 3396	.	.
	#M ADENOCARCINOMA	16	7	6	4	0.8853	0 5239	0.6342	0.8258
	C_FIBROADENOMA_M+B	59	25	28	22	0.8726	0 8388	0.5432	0.8915
OVARIES	#B GRANULOSA CELL TU	0	1	1	3	0.0129*	0 3396	0.3269	0.0309*
	#M ADENOCARCINOMA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	#M THECOMA, MALIGNAN	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	C_GRANULOSA_THECOMA_	1	1	1	3	0.0401	0 5653	0.5484	0.0953
PALATE	#B PAPILLOMA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
PANCREAS	#B ADENOMA, ACINAR C	0	0	0	1	0.1892	.	.	0.3182
	#B ADENOMA, ISLET CE	3	1	5	1	0.4593	0.4187	0.0761	0.3802
	#M CARCINOMA, ISLET	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	C_ISLET_CELL_M+B	4	1	5	1	0.5452	0 5537	0.1284	0.5107
PARATHYROIDS	#B ADENOMA	3	1	2	1	0.4988	0.4149	0.5242	0.3765
PAWS	#B OSTEOMA	0	0	0	1	0.1892	.	.	0.3182
PITUITARY	#B ADENOMA, PARS DIS	95	53	46	55	0.0725	0.1503	0.5781	0.0528
	#M CARCINOMA, PARS D	2	1	2	1	0.4310	0 2663	0.4036	0.6859
	C_PARS_DISTALIS_M+B	97	54	48	56	0.0608	0.1429	0.4794	0.0466
SKIN	#B F BROMA	0	1	0	0	0.3861	0 3396	.	.
	#B LIPOMA	0	0	1	0	0.3861	.	0.3269	.
	#M CARCINOMA	0	0	2	0	0.3813	.	0.1055	.
	#M F BROSARCOMA	1	1	0	4	0.0105	0 5653	0.3269	0.0377
	C_FIBROMA_M+B	1	2	0	4	0.0200	0 2663	0.3269	0.0377
SKIN/SUBCUTIS	C_FIBROSARCOMA_M	1	2	1	4	0.0217	0 2719	0.5484	0.0377
	C_FIBROSARCOMA_M+B	1	3	2	4	0.0396	0.1177	0.2550	0.0377
SOFT TISSUE- AB	#B LIPOMA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	#M HIBERNOMA, MALIGN	0	1	0	0	0.3861	0 3396	.	.
SOFT TISSUE- TH	#B HIBERNOMA, BENIGN	0	1	1	1	0.1617	0 3396	0.3269	0.3182
STOMACH NON	C_PAP LLOMA_SQUAMOUS	0	1	1	0	0.3799	0 3438	0.3269	.
STOMACH, NON	#B PAPILLOMA	0	0	1	0	0.3861	.	0.3269	.
	#B PAPILLOMA, SQUAMO	0	1	0	0	0.3846	0 3438	.	.
SUBCUTIS	#B ADENOMA	0	0	1	0	0.3861	.	0.3269	.
	#B F BROMA	0	0	1	0	0.3885	.	0.3312	.
	#B LIPOMA	1	0	0	1	0.3432	0 3396	0.3269	0.5365
	#M F BROSARCOMA	0	1	1	0	0.3799	0 3438	0.3269	.
	#M SARCOMA, SP NDLE	0	0	0	1	0.1892	.	.	0.3182
SYSTEMIC TUMORS	#M HEMANGIOSARCOMA	0	1	0	0	0.3861	0 3396	.	.
	#M LEUKEMIA	1	0	1	1	0.2534	0 3396	0.5484	0.5365
	#M LYMPHOMA, MALIGNA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	#M MESOTHELIOMA, MAL	0	0	0	1	0.1892	.	.	0.3182
	#M SARCOMA, HISTIOCY	2	3	0	1	0.6338	0 2225	0.5484	0.6859
TA L	#B OSTEOMA	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	#B SCHWANNOMA, BENIG	0	0	0	1	0.1892	.	.	0.3182
THYMUS	#M THYMOMA, MALIGNAN	1	0	1	0	0.4638	0 3396	0.5484	0.3182
THYROID GLANDS	#B ADENOMA, C-CELL	18	5	6	7	0.5826	0 8603	0.7414	0.5873
	#B ADENOMA, FOLLICUL	6	1	3	2	0.5320	0.7540	0.6229	0.4960
	#M CARCINOMA, C-CELL	1	4	1	0	0.8309	0 0458	0.5484	0.3182

File Name: outputfile792527110.pdf

Organ Name	Tumor Name	0 mg/kg/day C N=130	2 mg/kg/day LD N=65	7 mg/kg/day MD N=65	20 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
	#M CARCINOMA, FOLLIC	1	0	0	0	0.5946	0 3396	0.3269	0.3182
	C_C_CELL_M+B	19	9	7	7	0.7344	0.4849	0.6800	0.6430
	C_FOLLICULAR_CELL_M+	7	1	3	2	0.6159	0 8220	0.4347	0.5895
UR NARY BLADDER	#B PAPILLOMA	0	0	1	0	0.3861	.	0.3269	.
UTERUS	#B LEIOMYOMA	0	0	2	0	0.3813	.	0.1055	.
	#B LIPOMA	0	0	1	0	0.3861	.	0.3269	.
	#B POLYP	4	10	5	5	0.2678	0 0036 *	0.1349	0.1158
	#M ADENOCARCINOMA	0	1	0	0	0.3861	0 3396	.	.
	#M CARCINOMA	2	0	0	0	0.8366	0 5653	0.5484	0.5365
	#M SARCOMA, ENDOMETR	0	1	0	0	0.3861	0 3396	.	.
	#M SCHWANNOMA, MALIG	0	1	0	0	0.3846	0 3438	.	.
	C_CARC NOMA_ADEN_M	2	1	0	0	0.8315	0 2663	0.5484	0.5365
VAG NA	#B GRANULAR CELL TUM	0	0	0	3	0.0064*	.	.	0.0309*
	#B POLYP	0	1	0	0	0.3861	0 3396	.	.
	#M CARCINOMA	0	0	1	0	0.3885	.	0.3312	.
ZYMBAL'S GLANDS	#M CARCINOMA, SQUAMO	1	0	2	1	0.2234	0 3396	0.2550	0.5365
Multi/Organ	C_PAP LLOMA_B	1	1	3	0	0.5812	0 5708	0.1026	0.3182

Table 6A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Male Mice

Organ Name	Tumor Name	0 mg/kg/day C N=130	4 mg/kg/day LD N=65	10 mg/kg/day MD N=65	25 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
ADRENAL CORTEX	#B ADENOMA	2	0	0	0	0.8452	0.5572	0.5572	0 5763
	#B ADENOMA, SUBCAPSU	1	0	0	0	0.6026	0.3309	0.3309	0 3453
COAGULATING GLA	#B ADENOMA	0	1	1	0	0.4157	0.3333	0.3333	.
DUODENUM	#M ADENOCARCINOMA	0	0	0	1	0.2105	.	.	0 3478
EPIDIDYMIDES	#M SARCOMA, HISTIOCY	1	0	0	0	0.6053	0.3333	0.3333	0 3478
HARDERIAN GLAND	#B ADENOMA	11	6	8	4	0.6891	0.4907	0.2508	0 6197
	#M CARCINOMA	0	2	0	0	0.6484	0.1127	.	.
JEJUNUM	#M ADENOCARCINOMA	0	0	0	1	0.2105	.	.	0 3478
KIDNEYS	#B ADENOMA	1	0	1	1	0.2870	0.3333	0.5572	0 5763
LIVER	#B ADENOMA, HEPATOCE	9	7	5	5	0.5195	0.2513	0.5447	0 5604
	#M CARCINOMA, HEPATO	9	2	6	6	0.2158	0.7714	0.3855	0.4175
	C_HEPATOCELLULAR_B+M	17	9	10	10	0.3641	0.5240	0.4498	0.4498
LUNGS	#B ADENOMA, BRONCHIO	19	6	6	9	0.5567	0.8035	0.8182	0 5120
	#M CARCINOMA, BRONCH	7	7	2	6	0.3058	0.1515	0.6351	0 2760
	C_BRONCHIOLO-ALVEOLA	22	12	8	14	0.3123	0.5214	0.7608	0 3412
PAWS	#B F BROMA	1	0	0	0	0.6053	0.3333	0.3333	0 3478
PHARYNX	#B F BROMA	1	0	0	0	0.6053	0.3333	0.3333	0 3478
PITUITARY	#M MENINGIOMA, MALIG	1	0	0	0	0.6053	0.3333	0.3333	0 3478
PROSTATE	#M ADENOCARCINOMA	0	0	1	0	0.4079	.	0.3333	.
SKIN	#M F BROSARCOMA	3	0	1	0	0.8418	0.7037	0.4030	0.7226

File Name: outputfile792527110.pdf

| Organ Name	Tumor Name	0 mg/kg/day C N=130	4 mg/kg/day LD N=65	10 mg/kg/day MD N=65	25 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
SKIN/SUBCUTIS	C_FIBROSARCOMA_M+B	4	0	1	3	0.2182	0.8010	0.5318	0.4627
STOMACH, NON	#B PAPILLOMA, SQUAMO	0	1	0	1	0.2105	0.3333	.	0 3478
SUBCUTIS	#M F BROSARCOMA	1	0	0	3	0.0312	0.3309	0.3309	0.1232
	#M LEIOMYOSARCOMA	1	0	0	1	0.3760	0.3309	0.3309	0 5730
	#M SARCOMA	1	0	0	0	0.6053	0.3333	0.3333	0 3478
SYSTEMIC TUMORS	#B HEMANGIOMA	2	3	3	0	0.7954	0.2061	0.2135	0 5763
	#M HEMANGIOSARCOMA	11	4	2	5	0.6267	0.6043	0.8737	0 5043
	#M LEUKEMIA, GRANULO	1	0	0	0	0.6026	0.3309	0.3309	0 3453
	#M LYMPHOMA, MALIGNA	13	2	6	6	0.4792	0.9238	0.4661	0 5025
	#M LYMPHOMA, PLASMA	1	0	0	0	0.6053	0.3333	0.3333	0 3478
	#M MESOTHELIOMA, MAL	0	0	0	1	0.2105	.	.	0 3478
	#M SARCOMA, HISTIOCY	1	3	3	1	0.4449	0.1100	0.1100	0 5730
TESTES	#B ADENOMA, LEYDIG C	2	0	2	1	0.3978	0.5572	0.4074	0 2773
	#M CARCINOMA, LEYDIG	0	0	1	0	0.4079	.	0.3333	.
THYROID GLANDS	#B ADENOMA, FOLLICUL	2	1	1	0	0.7585	0.7070	0.7070	0 5763

Table 6B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Female Mice

Organ Name	Tumor Name	0 mg/kg/day C N=130	4 mg/kg/day LD N=65	10 mg/kg/day MD N=65	25 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
ADRENAL CORTEX	#B ADENOMA	0	1	0	0	0.4219	0.3212	.	.
	#B ADENOMA, SUBCAPSU	2	1	0	1	0 5234	0.6872	0.5571	0.2949
ADRENAL MEDULLA	#B PHEOCHROMOCYTOMA,	0	0	1	0	0.4219	.	0.3357	.
	#M NEUROBLASTOMA	0	0	1	0	0.4219	.	0.3357	.
CECUM	#B LEIOMYOMA	0	0	0	1	0 2236	.	.	0.3630
CERVIX	#B GRANULAR CELL TUM	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#B LEIOMYOMA	0	2	0	1	0 3313	0.1015	.	0.3630
	#B POLYP, ENDOMETRIA	0	0	0	2	0 0493	.	.	0.1302
	#M CARCINOMA, SQUAMO	0	1	0	0	0.4219	0.3212	.	.
	#M F BROSARCOMA	0	1	0	0	0.4219	0.3212	.	.
	#M GRANULAR CELL TUM	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#M LEIOMYOSARCOMA	0	1	0	1	0 2217	0.3212	.	0.3630
	#M SARCOMA, ENDOMETR	0	1	0	0	0.4219	0.3212	.	.
	#M SCHWANNOMA, MALIG	0	0	1	0	0.4219	.	0.3357	.
CERVIX/VAG NA	C_GRANULAR_CELL_B+M	2	1	0	0	0 8529	0.6872	0.5571	0.5927
CERVIX/UTERUS	C_POLYP_ENDOMETRIAL_	10	3	4	5	0 5455	0.6457	0.5442	0.4856
FEMUR	#B OSTEOMA	1	0	0	0	0 6050	0.3188	0.3333	0.3605
HARDERIAN GLAND	#B ADENOMA	2	0	3	4	0 0341	0.5408	0.2170	0.1272
	#M CARCINOMA	1	0	1	0	0 5106	0.3212	0.5603	0.3630
HARDERIAN/GLAND	C_ADENOMA+CARCI	3	0	4	4	0 0673	0.6905	0.1788	0.2168
ILEUM	#M ADENOCARCINOMA	0	1	0	0	0.4219	0.3212	.	.
LIVER	#B ADENOMA, HEPATOCE	6	3	2	3	0 5829	0.5979	0.5408	0.4230
	#M CARCINOMA, HEPATO	1	0	1	4	0.0119*	0.3212	0.5603	0.0584
	C_HEPATOCELLULAR_B+M	7	3	3	7	0.1248	0.4063	0.4474	0.2125
LN, ILIAC	#M SARCOMA, UND FFER	0	0	0	1	0 2269	.	.	0.3673

File Name: outputfile792527110.pdf

Organ Name	Tumor Name	0 mg/kg/day C N=130	4 mg/kg/day LD N=65	10 mg/kg/day MD N=65	25 mg/kg/day HD N=65	P-Value Dose Response	C vs. LD	C vs. MD	C vs. HD
LUNGS	#B ADENOMA, BRONCHIO	14	4	8	7	0 5091	0.7296	0.4797	0.5050
	#M CARCINOMA, BRONCH	7	3	2	5	0 3619	0.4202	0.6465	0.4807
	C_BRONCHIOLO_ALVEOLA	19	7	10	12	0 3148	0.6284	0.5348	0.4484
MAMMARY GLAND	#M ADENOCARCINOMA	1	0	1	0	0 5114	0.3212	0.5666	0.3630
	#M ADENOCARCINOMA, M	0	1	0	0	0.4219	0.3212	.	.
OVARIES	#B ADENOMA, TUBULOST	1	1	1	0	0 6291	0.5408	0.5603	0.3630
	#B CYSTADENOMA	3	0	1	1	0 5627	0.6905	0.4201	0.4606
	#B GRANULOSA CELL TU	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#B LUTEOMA	2	2	1	1	0 5677	0.3855	0.2608	0.2983
	#B LUTEOMA, MULT PLE	0	0	1	0	0.4219	.	0.3357	.
	#B SERTOLI CELL TUMO	0	0	0	1	0 2236	.	.	0.3630
	#B THECOMA, BENIGN	1	0	0	1	0.4016	0.3188	0.3333	0.5982
	#M CHORIOCARCINOMA	0	0	0	1	0 2236	.	.	0.3630
	#M SARCOMA	1	0	0	0	0 6076	0.3212	0.3357	0.3630
PITUITARY	#B ADENOMA, PARS NT	0	0	0	1	0 2236	.	.	0.3630
RECTUM	#M CARCINOMA, SQUAMO	0	0	0	1	0 2236	.	.	0.3630
SKIN	#B PAPILLOMA, SQUAMO	0	0	0	1	0 2269	.	.	0.3673
	#B TUMOR, HA R FOLLI	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#M F BROSARCOMA	1	0	1	2	0.1202	0.3212	0.5666	0.3044
	#M HIBERNOMA, MALIGN	0	1	0	0	0.4202	0.3261	.	.
	#M SARCOMA, UND FFER	0	1	1	0	0.4365	0.3212	0.3404	.
SKIN/SUBCUTIS	C_FIBROSARCOMA_M+B	4	0	2	2	0.4405	0.7892	0.3267	0.3819
SOFT TISSUE- AB	#M OSTEOSARCOMA	0	1	0	0	0.4219	0.3212	.	.
SUBCUTIS	#B LEIOMYOMA	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#M F BROSARCOMA	3	0	1	0	0 8513	0.6872	0.4074	0.7416
SYSTEMIC	C_HEMANGIOMA_M+B	11	9	7	3	0 9114	0.1424	0.4131	0.8105
SYSTEMIC TUMORS	#B HEMANGIOMA	3	0	3	0	0.7794	0.6905	0.3315	0.7446
	#M HEMANGIOSARCOMA	8	9	4	3	0 8472	0.0493	0.3718	0.6023
	#M LYMPHOMA, MALIGNA	39	13	21	16	0.7631	0.8593	0.3958	0.8069
	#M SARCOMA, HISTIOCY	9	4	7	5	0.4833	0.4094	0.2869	0.4078
THYROID GLANDS	#B ADENOMA, FOLLICUL	0	1	0	0	0.4219	0.3212	.	.
TONGUE	#B PAPILLOMA, SQUAMO	1	0	0	0	0 6076	0.3212	0.3357	0.3630
UTERUS	#B F BROMA	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#B GRANULAR CELL TUM	0	1	0	0	0.4219	0.3212	.	.
	#B LEIOMYOMA	0	2	0	0	0 6669	0.1015	.	.
	#B POLYP, ENDOMETRIA	10	3	4	3	0 8260	0.6457	0.5442	0.7694
	#M ADENOCARCINOMA	2	0	0	0	0 8470	0.5408	0.5603	0.5958
	#M F BROSARCOMA	0	0	1	0	0.4219	.	0.3357	.
	#M LEIOMYOSARCOMA	0	1	1	0	0.4366	0.3212	0.3357	.
	#M SCHWANNOMA, MALIG	2	0	0	0	0 8470	0.5408	0.5603	0.5958
VAG NA	#B F BROMA	1	0	0	0	0 6076	0.3212	0.3357	0.3630
	#M CARCINOMA, SQUAMO	1	1	0	0	0 6999	0.5474	0.3357	0.3630
	#M GRANULAR CELL TUM	0	1	0	0	0.4219	0.3212	.	.
Multi/Organ	C_PAP LLOMA_B	1	0	0	1	0.4030	0.3212	0.3357	0.6013

File Name: outputfile792527110.pdf

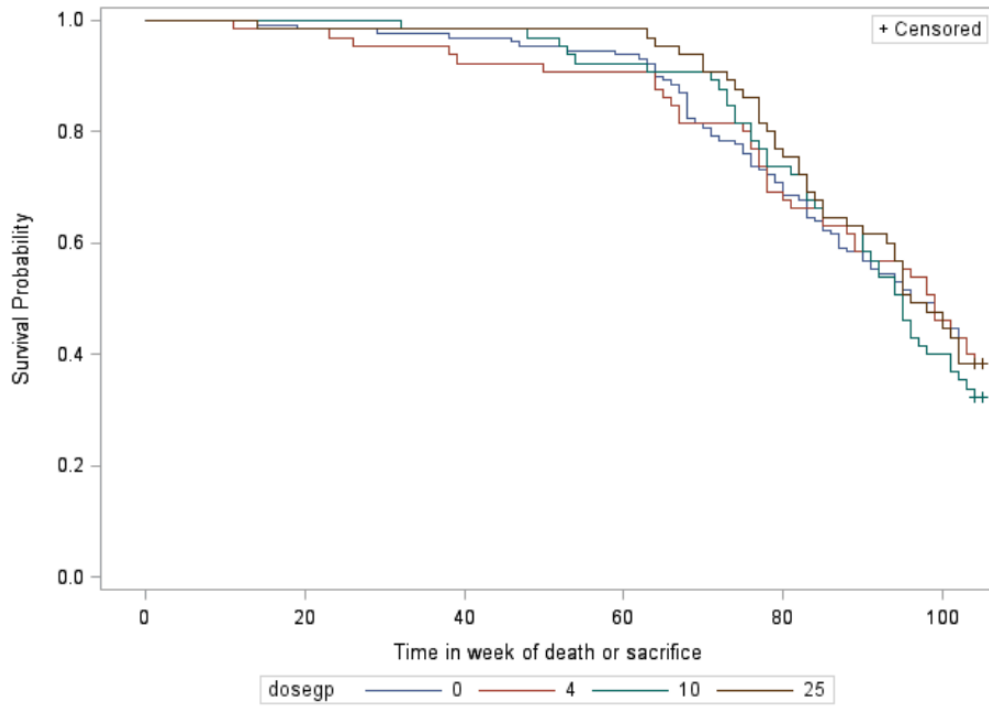**Figure 1A: Kaplan-Meier Survival Functions for Male Rats**
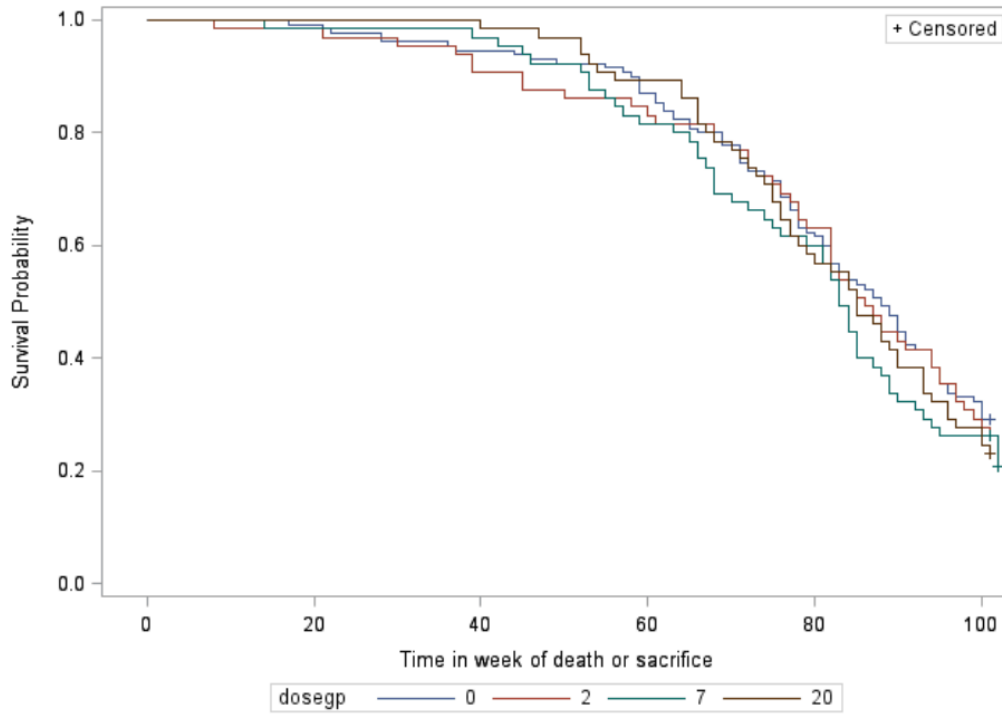
Note: dose group should be 0, 2, 7, or 20-mg/kg/day

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

Note: dose group should be 0, 2, 7, or 20-mg/kg/day

File Name: outputfile792527110.pdf
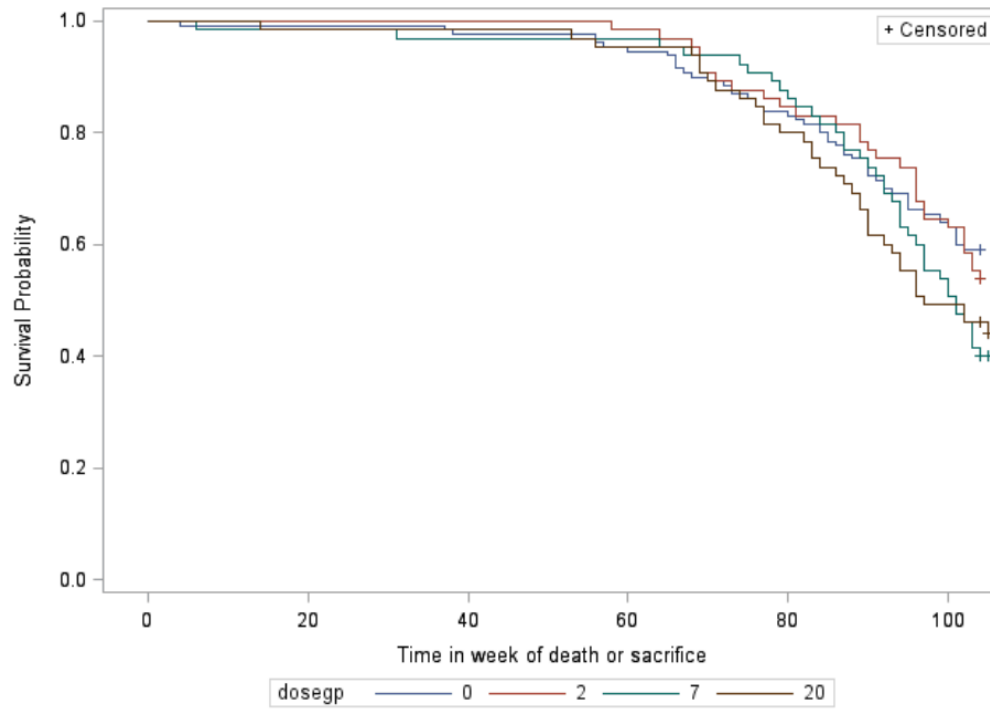
Figure 2A: Kaplan-Meier Survival Functions for Male Mice

Note: dose group should be 0, 4, 10, or 25-mg/kg/day

Figure 2B: Kaplan-Meier Survival Functions for Female Mice

Note: dose group should be 0, 4, 10, or 25-mg/kg/day

File Name: outputfile792527110.pdf
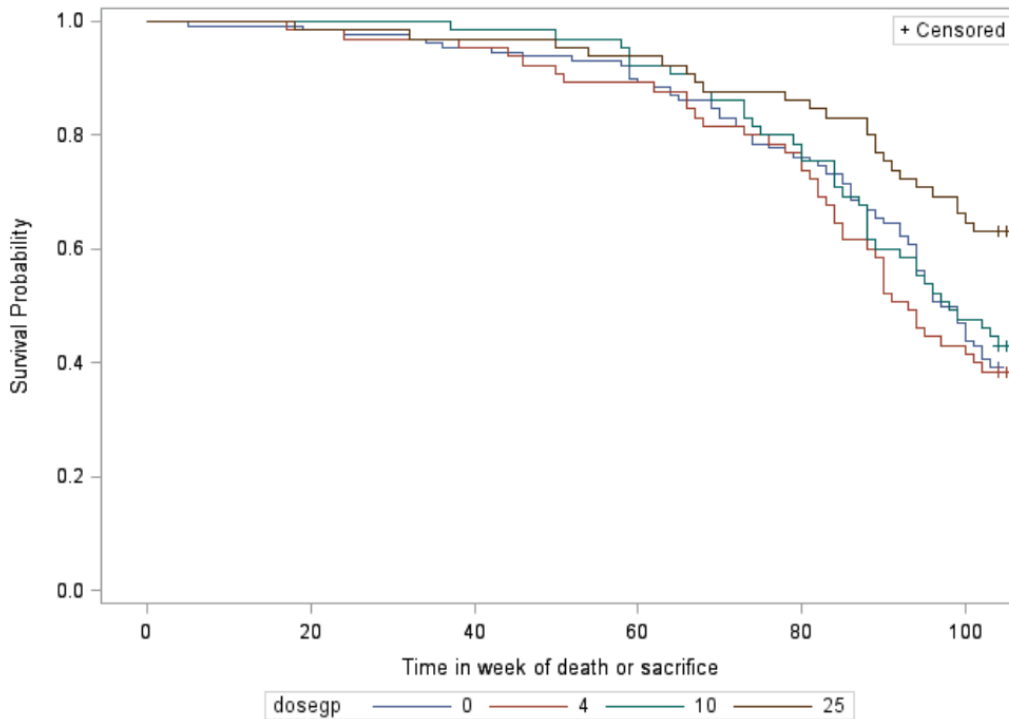
7   References

1. Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J.Wahrendorf, "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", <u>Long term and short term screening assays for carcinogens: A critical appraisal</u>, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva,* 311-426, 1980.
2. 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.
3. Bieler, G. S. and Williams, R. L. (1993). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity". *Biometrics* 49, 793-801.
4. Tarone RE, "Test for trend in life table analysis", *Biometrika* 1975, 62: 679-82
5. Lin K.K. and Rahman M.A.," 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, 1998.
6. Haseman, J, "A re-examination of false-positive rates for carcinogenesis studies", *Fundamental and Applied Toxicology*, 3: 334-339, 1983.
7. Guidance for Industry. Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Statues of Pharmaceuticals (Draft Guidance). U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), May 2001.

--EOF--

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

/s/

FENG ZHOU
09/09/2015

KARL K LIN
09/09/2015
Concur with review