CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

208692Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 208692
Drug Name:	Cabometyx (cabozantinib)
Indication(s):	(b) (4) Renal Cell Carcinoma (b) (4) Prior (b) (4) (b) (4) Kinase Inhibitor Therapy
Applicant:	Exelexis
Date(s):	Submission Date: 22 December 2015 PDUFA Date: 13 May 2016
Review Priority:	Priority
Biometrics Division:	Division of Biometrics V
Statistical Reviewer:	Joyce Cheng
Concurring Reviewers:	Shenghui Tang, Team Leader Rajeshwari Sridhara, Division Director
Medical Division:	OHOP/DOP1
Clinical Team:	Michael Brave, Harpreet Singh Julia Beaver, Team Leader
Project Manager:	Rajesh Venugopal
Keywords:	

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Cabozantinib is an oral small molecule inhibitor of tyrosine kinases including MET, VEGFRs, and AXL. It is currently approved for the treatment of patients with progressive metastatic medullary thyroid cancer. In this NDA, Exelixis seeks approval of cabozantinib for the treatment of advanced renal cell carcinoma (RCC) in patients who have received (b) (4) prior therapy.

Efficacy for this proposed indication was based on results from METEOR, a randomized, controlled, Phase 3 study in subjects with advanced RCC who had progressed after at least one prior VEGFR-TKI. Eligible subjects were randomized 1:1 to receive cabozantinib or everolimus. The primary endpoint was progression-free survival (PFS) with key secondary endpoints of overall survival (OS) and objective response rate (ORR). The study was powered for both PFS and OS and recruited 658 patients (Intent-to-Treat (ITT) population), but the primary PFS analysis was conducted in the first 375 subjects randomized (Primary Endpoint Intent-to-Treat (PITT) population) which was pre-specified and agreed upon by the FDA.

The primary PFS analysis showed a statistically significant 3.6 month improvement in median PFS for cabozantinib over everolimus with a stratified hazard ratio of 0.58 (95% CI: 0.45, 0.74) in the PITT population with stratification data from the IVRS at the time of randomization. A sensitivity analysis for PFS performed using the full ITT population showed a statistically significant 3.5 month improvement in median PFS for cabozantinib over everolimus with a stratified hazard ratio of 0.51 (95% CI: 0.41, 0.62). The pre-specified interim analysis of OS was performed at the time of the primary PFS analysis with 49% information and showed longer survival in the cabozantinib arm with a stratified hazard ratio of 0.68 (95% CI: 0.52, 0.90), but the critical p-value necessary to determine significance at interim was not reached. A second unplanned interim analysis was performed, at which point 78% information was available, and resulted in a statistically significant 4.9 month improvement in median OS in the cabozantinib arm with a stratified hazard ratio of 0.66 (95% CI: 0.53, 0.83). Additionally, there was a statistically significant improvement in ORR for the cabozantinib arm over the everolimus arm (17% vs. 3% confirmed partial responses).

A number of sensitivity analyses performed by the applicant as well as the FDA support the robustness of the primary PFS findings. Based on the evidence from the METEOR study, cabozantinib appears to have a PFS, OS, and ORR benefit over everolimus. The final decision and benefit-risk evaluation of whether the magnitude of this benefit is clinically meaningful with an acceptable safety profile is deferred to the clinical review team.

2 INTRODUCTION

2.1 Overview

Cabozantinib inhibits multiple receptor tyrosine kinases (RTKs) implicated in angiogenesis, invasion, or metastasis in renal cell carcinoma (RCC), including MET, VEGFRs, and AXL. In this New Drug Application (NDA), Exelexis seeks approval of cabozantinib for the indication of treatment of advanced RCC in patients who have received (b) (4) prior therapy.

2.1.1 Background

Renal cell carcinoma (RCC) is the third most frequent cancer of the urinary tract. Kidney cancer is diagnosed in about 330,000 individuals worldwide each year and results in 140,000 deaths annually. The incidence of RCC has generally increased over the past several years. There is a higher incidence and mortality in men than woman. Incidence peaks between ages 60 and 70.

The current standard of care for advanced RCC patients whose disease has progressed on or who are resistant to VEGFR-TKI therapy is treatment with everolimus or axitinib. Sorafenib is also recommended in this setting based on a study in patients who had progressed on a prior systemic (mainly cytokine-based) therapy. Everolimus is the most frequently used second-line therapy following a VEGFR-TKI in patients with RCC. However, these approved second-line options for patients (everolimus and axitinib) are limited by modest PFS benefit and a lack of improvement in OS. In November 2015, nivolumab was approved based on improved OS but no PFS benefit as a second-line treatment.

As reported by the Sponsor, cabozantinib inhibits several RTKs known to influence tumor growth, metastasis, and angiogenesis, including MET, VEGFRs, and AXL. Based on the molecular pathobiology of RCC, there is a strong mechanistic rationale for the evaluation of cabozantinib in this disease.

2.1.2 Regulatory History and Changes in Study Conduct or Planned Analyses

Cabozantinib is currently approved for the treatment of patients with progressive, metastatic medullary thyroid cancer. In January 2013, the applicant discussed trial design for the RCC indication with FDA and four national regulatory agencies in Europe. Cabozantinib has been granted fast track designation (April 2015) and breakthrough therapy designation (August 2015) for the treatment of advanced RCC in patients who have received one prior therapy.

There was one amendment to the original protocol as well as several country-specific amendments. There were three versions of the statistical analysis plan (SAP). The second version was completed before analysis of the primary endpoint. The third version was completed after analysis of the primary endpoint and contains minor changes in operational conventions that were adopted prior to analysis of the primary endpoint and minor editorial changes. It also includes a second unplanned interim analysis of OS conducted with a data cutoff of 31 December 2015. Results from this analysis were provided in an OS Addendum (received 11 February 2015) to the clinical study report (CSR).

Additionally, a Per Protocol population was described in the SAP but was never specifically defined, and no analyses using that population were reported in the CSR. Also, the algorithm used to convert KPS to ECOG PS in the CSR was different from the one presented in the SAP.

2.1.3 Specific Studies Reviewed

Support for this indication comes from Study XL184-308 (METEOR), a randomized, controlled, Phase 3 study in subjects with advanced RCC who had progressed after at least one prior VEGFR-TKI. The study was initiated after antitumor activity was observed in a Phase 1 study that enrolled heavily pretreated subjects with RCC. In the Phase 3 study, everolimus was used as the comparator and results showed that cabozantinib demonstrated a statistically significant improvement in PFS, OS (at the second unplanned interim OS analysis), and ORR. This review will primarily be based on the Phase 3 study. An overview is given in Table 1.

Table 1: Overview of XL184-308 (METEOR)

Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
Phase 3, multicenter, randomized, open-label, controlled	Subjects who met all study eligibility criteria were randomly assigned 1:1 to open-label treatment with either cabozantinib or everolimus. Randomization was stratified. Subjects received study treatment as long as they continued to experience clinical benefit by investigator opinion or until unacceptable toxicity, need for subsequent systemic anticancer treatment, or any other reason as listed in protocol. Treatment was allowed to continue after radiographic RCC progression per RECIST 1.1 if the investigator believed that the subject was still receiving clinical benefit from study treatment and that potential benefit of continuing study treatment outweighed the potential risks. Crossover between arms was not allowed.	Thirty days (+14) after the date of the decision to permanently discontinue study treatment subjects returned for a Post-Treatment Follow-up Visit for safety assessments. Radiographic tumor assessments and HRQOL assessments were continued, regardless of whether study treatment was given, reduced, held, or discontinued until the criteria for discontinuation of imaging assessments were met. Consequently these assessments were required in the Post Treatment Period for some subjects. Subjects were contacted every 8 weeks (±7 days) after the Post-Treatment Follow-up Visit to assess survival status and document receipt of subsequent anticancer therapy. These assessments were continued until the subject expired or Exelixis decided to discontinue collection of these data in the study.	ITT Population: Cabozantinib 60 mg (n=330) Everolimus (n=328) PITT Population (first 375 randomized subjects): Cabozantinib 60 mg (n=187) Everolimus (n=188)	Required to have radiographic progression within 6 months after last dose of study treatment and to have received at least one prior VEGFR-TKI. Other key entry criteria included: histologically or cytologically confirmed advanced RCC with a clear cell component, ≥ 18 years of age, Karnofsky performance status ≥ 70, measurable disease by CT/MRI per RECIST 1.1, and rPD during VEGFR-TKI treatment or treated for ≥ 4 weeks with rPD within 6 months after last dose of VEGFR-TKI

2.2 Data Sources

The electronic submission, including protocols, statistical analysis plan, study reports, and analysis datasets, for this NDA can be accessed here:
\\CDSESUB1\evsprod\NDA208692\208692.enx

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The data and analysis quality of the submission was acceptable for the reviewer to perform the statistical review.

3.2 Evaluation of Efficacy

The efficacy evaluation focuses on the METEOR study for the proposed indication in RCC.

3.2.1 Study Design and Endpoints

METEOR was a Phase 3, multicenter, randomized, open-label, controlled study to evaluate the effect of cabozantinib compared with everolimus on PFS and OS in subjects with advanced RCC that had progressed after prior VEGFR TKI therapy. Open-label design enabled appropriate dose modifications for AEs in both treatment arms and, in order to prevent bias, the independent radiology committee (IRC) was blinded to treatment identity and clinical data that may lead to inadvertent unblinding.

Eligible subjects were randomized 1:1 to receive cabozantinib or everolimus with a target enrollment of 650 subjects in 25 countries at 173 sites in four regions (principally North America and Europe, but also Asia Pacific/Australia and Latin America). Treatment regimens were as follows:

1. Oral cabozantinib (60 mg) tablets once-daily qd
2. Oral everolimus (10 mg) tablets qd

Randomization was stratified by two factors:

1. Number of prior VEGFR-targeting TKI therapies (1 vs 2 or more)
2. Number of risk factors per Memorial Sloan-Kettering Cancer Center (MSKCC) prognostic criteria for previously treated patients with RCC (0 vs 1 vs 2 or 3). The three risk factors were:
    a. KPS < 80%
    b. Hgb < 13 g/dL (130 g/L) for males and < 11.5 g/dL (<115 g/L) for females
    c. Corrected serum calcium > upper limit of normal (ULN)

Subjects received study treatment as long as they continued to experience clinical benefit, as determined by the investigator. Reasons for discontinuation of treatment included, among others, an unacceptable toxicity or the need for subsequent systemic anti-cancer therapy. Crossover between treatment arms was not allowed to maximize the ability to evaluate the effect of cabozantinib on OS.

Clinic visits occurred at regular intervals through a 30-day post-treatment follow-up visit, with contact every 8 weeks thereafter to assess survival status and to document receipt of subsequent anticancer therapy. Both PFS and ORR were based on RECIST 1.1 per IRC. CT (or MRI) of chest/abdomen/pelvis (CAP) was performed in all subjects at screening, every 8 weeks (± 5 days) after randomization throughout the first 12 months on study, and then every 12 weeks afterwards (± 7 days). Further details are given in Appendix A.

The primary and secondary efficacy endpoints for METEOR are listed below:

Primary endpoints:

- PFS per RECIST 1.1 per IRC, defined as the time from randomization to the earlier of the following events: documented progressive disease (PD) per RECIST 1.1 or death due to any cause.

Key secondary endpoints:

- OS, defined as the time from first dose until death due to any cause. If a subject withdrew consent to participate in the study, no further study data were collected for this subject, other than the determination of survival status from public records such as government vital statistics or obituaries.
- ORR per RECIST 1.1 per IRC, defined as the proportion of subjects for whom the best overall response at the time of data cutoff was complete response (CR) or partial response (PR) as assessed by the IRC per RECIST 1.1, which was confirmed by a subsequent visit $\geq$ 28 days later.

Additional exploratory endpoints:

- Duration of response (DOR)
- Changes in bone scans
- Safety and tolerability
- Characterization of the pharmacokinetics of cabozantinib
- Change in kidney-cancer related symptoms as assessed by the Functional Assessment of Cancer Therapy-Kidney Cancer Symptom Index (FKSI-19)
- Change in mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, and global health as assessed by the EuroQol Health questionnaire instrument (EQ-5D-5L)
- Proportion of subjects with post-randomization skeletal-related events (SREs)
- Relationship of baseline and changes in plasma biomarkers, serum bone markers, serum calcium, and circulating tumor cells (CTCs) with treatment and/or clinical outcome (not analyzed for this report)
- Health care resource utilization

The METEOR study used a design, depicted in Figure 1, to provide adequate power to evaluate both PFS and OS, even though PFS was the primary endpoint. For PFS, 259 events were required detect a hazard ratio of 0.667 (corresponding to an increase from 5 to 7.5 months in median PFS) with 90% power using a log-rank test at a two-sided significance level of 5%. It was estimated that 375 subjects would be adequate to observe 259 events to evaluate the primary PFS endpoint. For the key secondary efficacy endpoint of OS, 408 deaths were required to detect a hazard ratio of 0.75 (corresponding to an increase from 15 to 20 months in median OS) with 80% power using a log-rank test at a two-sided significance level of 4%.

Assuming an average accrual rate of 32 subjects per month and a 1:1 treatment allocation ratio, Exelixis determined a total of 650 subjects (325 per treatment arm) would be needed to observe

the required number of events within the planned study duration, which consisted of 21 months accrual, then approximately 17 months to observe the required PFS events among 375 subjects and approximately 36 months to observe the required deaths for OS among 650 subjects. Note that the number of events necessary to evaluate the primary PFS endpoint could have occurred before the study was fully accrued. In order to not bias the results towards subjects whose disease progressed early, the primary analysis of PFS was determined from only the first 375 subjects randomized.

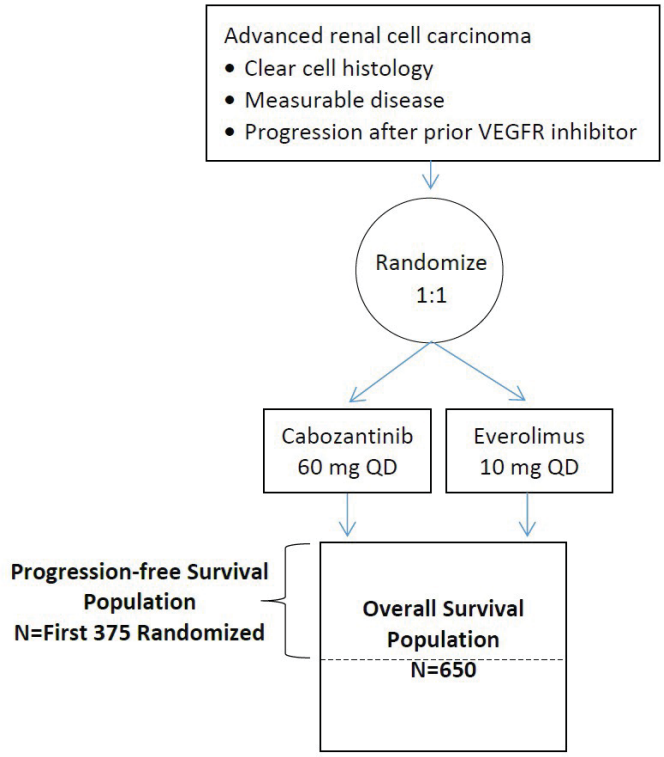


Figure 1: METEOR Study Design
[Source: CSR Figure 3]

There was a pre-specified interim analysis conducted for the key secondary efficacy endpoint OS at the time of the primary analysis of PFS. At that time, 202 (49%) deaths were observed out of the 408 required for final OS evaluation. Type 1 error was controlled by a Lan-DeMets O'Brien-Fleming alpha spending function to account for the actual information at the time of interim analysis (critical value 0.0019). A second unplanned interim analysis was conducted with a data cutoff of 31 December 2015, providing a minimum of 12 months of follow-up. The critical value for the second OS interim analysis was calculated based on the actual observed number of deaths. The protocol included no criteria for futility stopping.

The study is ongoing, but study endpoints have been analyzed, including an interim OS analysis using a data cutoff of 22 May 2015 and a second unplanned interim analysis using a data cutoff of 31 December 2015 with results provided in an addendum to the CSR. Subjects continue to be followed for OS and safety.

3.2.2 Statistical Methodologies

Due to the study design, the applicant defined two efficacy analysis populations. The Primary Endpoint Intent-to-Treat (PITT) population consisted of the first 375 randomized subjects (187 in the cabozantinib arm and 188 in the everolimus arm) and was used to determine the primary endpoint PFS of the study. The Intent-to-Treat (ITT) population, defined as all randomized subjects, had a total of 658 subjects (330 in the cabozantinib arm and 238 in the everolimus arm) and was used for efficacy analyses other than the primary analysis of PFS.

The primary analysis of the primary PFS performed on the PITT population was designed to only include progression events per IRC per RECIST 1.1, not including clinical deterioration or radiographic progression determined by investigator. The general censoring rules used are described below:

- Subjects who receive non-protocol anti-cancer therapy (NPACT) after randomization (including medications (systemic or local) or radiation to soft tissue) before experiencing an event will be right censored at the date of the last adequate tumor assessment prior to the date of initiation of subsequent therapy.
- Subjects who experience surgical resection of tumor lesions after randomization before experiencing an event will be right censored at the date of the last adequate tumor assessment prior to the date of the surgery.
- Subjects who have not experienced an event (and are not otherwise censored) at the time of data cutoff will be right censored on the date of their last adequate tumor assessment.
- Subjects who miss 2 or more consecutive adequate scheduled tumor assessments immediately followed by an event will be right censored on the date of their most-recent adequate tumor assessment prior to the missing/inadequate assessments.
    - If the 2 or more consecutive missing adequate assessments are immediately followed by an adequate assessment with an overall response assignment of SD, PR, or CR, this will be deemed sufficient clinical evidence that progression did not occur during the period of missing data and the missing evaluations will be ignored.

Only adequate tumor assessments (ATAs), defined as an evaluation performed per RECIST 1.1 resulting in an overall response of CR, PR, SD, or PD, were considered in the determination of progression and censoring dates. Unless PD was otherwise evident, partially missing tumor data or indeterminate lesions for a particular tumor assessment resulted in an overall response of "not evaluable" and the tumor assessment was not considered adequate. Single missing or inadequate scheduled tumor assessments were ignored. No values were imputed.

After 259 events were observed in the first 375 patients and the enrollment of 650 subjects was completed, PFS was descriptively summarized using the Kaplan-Meier method and compared between the treatment arms using a stratified log-rank test with a 2-sided 0.05 level of significance. The hazard ratio was estimated using a stratified Cox regression model and stratification factors were the same as those used for randomization. Planned supportive analyses for PFS included various sensitivity analyses based upon different definitions of progression events, censoring rules, and analysis populations.

Key secondary efficacy endpoints were OS and ORR. The final analysis of OS is planned to be conducted after study enrollment is complete and at least 408 deaths are observed in the study. Subjects who were alive at the time of data cutoff or were permanently lost to follow up were right censored at either the data cutoff date or the date the subject was last known to be alive, whichever was earlier. Statistical analysis of OS followed that of PFS. An interim analysis of OS was performed at the time of primary PFS analysis using the ITT population available at that time with type 1 error controlled by a Lan-DeMets O'Brien-Fleming alpha spending function. A second unplanned interim analysis was conducted with a data cutoff of 31 December 2015, providing a minimum of 12 months of follow-up.

The analysis of ORR per RECIST 1.1 was based upon assessments as determined by the IRC. Hypothesis testing was performed using the chi-squared test with a two-sided significance level of 0.01. If a sufficient number of responders were observed, the Cochran-Mantel-Haenszel method to adjust for randomization stratification factors was used in the analysis. Point estimates of ORR, difference in response rates between the two arms, and associated confidence intervals (calculated using exact methods) were also provided. Analysis of ORR was performed at the time of primary PFS analysis using the entire ITT population available at that time. Subjects with no post-baseline tumor assessments were counted as non-responders. A waterfall plot of best percentage change in target lesion size per IRC was presented for each treatment group using the ITT and PITT populations. Similar supportive analyses were also conducted using investigator assessed ORR.

The multiplicity issue resulting from analysis of one primary endpoint (PFS), two key secondary efficacy endpoints (ORR and OS), and performing one interim analysis (of OS) was addressed by employing a fixed-sequence testing procedure, applying a modified Bonferroni procedure (dividing the alpha between the key secondary endpoints: 0.04 for OS, 0.01 for ORR), and implementing an alpha spending function.

Exploratory efficacy endpoints were analyzed in both the PITT and ITT populations using appropriate 2-sided statistical tests without adjustment for multiplicity and results were considered descriptive.

Reviewer's Comments:

1. *The second unplanned interim analysis for OS was added in version 3 of the SAP submitted 06 October 2015. The pre-specified Lan-Demets O'Brien-Fleming alpha spending function can continue to be used to control type-1 error in this analysis, with the critical p-value dependent on the information fraction available.*
2. *The study ended up enrolling 658 subjects total. The applicant stated that the primary PFS endpoint (based on first 375 subjects enrolled) actually included 247 events because there is a delay between the date a radiographic progression event occurs and the date it is ascertained due to the time required for radiographic images and censoring data to be obtained from study sites, image QC and review that is performed by the IRC, and the application of censoring rules. It was predicted that the 259th event would occur by the cutoff date of 22 May 2015 but in fact only 247 events had been reached, which was judged to be sufficiently close to target.*

3. *Waterfall plots do not include patients with unmeasurable disease, non-target lesions, and new lesions.*

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

3.2.3.1 Patient Disposition

Patient disposition is summarized in Table 2. As of the data cutoff date of 22 May 2015, a lower rate of study treatment discontinuation was observed in the cabozantinib arm compared with the everolimus arm (60% vs 80%; ITT population). This difference was primarily due to the higher rate of disease progression (37% vs 48%) and clinical deterioration (8.5% vs 16%) in the everolimus arm. Other reasons for discontinuation were similar between arms; the rate of treatment discontinuation due to an AE was 10% in the cabozantinib arm and 9.5% in the everolimus arm.

Similar results were observed in the PITT population. As of the data cutoff date, 70% of subjects in the cabozantinib arm and 82% of subjects in the everolimus arm had discontinued study treatment. Disease progression was the primary reason for discontinuation and was higher in the everolimus arm (44% cabozantinib, 49% everolimus). Treatment discontinuation due to an AE was balanced between arms (11% cabozantinib, 11% everolimus).

Table 2: Summary of Patient Disposition

	PITT Population		ITT Population	
	Cabozantinib	**Everolimus**	**Cabozantinib**	**Everolimus**
	N = 187	**N = 188**	**N = 330**	**N = 328**
Never treated	0	3 (1.6)	0	5 (1.5)
Treatment ongoing	56 (30)	33 (18)	133 (40)	67 (20)
Discontinued treatment				
Progressive disease	82 (44)	92 (49)	122 (37)	158 (48)
Adverse event	21 (11)	20 (11)	32 (10)	31 (9.5)
Clinical deterioration	18 (10)	29 (15)	28 (8.5)	51 (16)
Withdrawal by subject	3 (1.6)	7 (3.7)	6 (1.8)	11 (3.4)
Physician decision	4 (2.1)	2 (1.1)	5 (1.5)	2 (0.6)
Lack of efficacy	2 (1.1)	0	3 (0.9)	0
Protocol violation	1 (0.5)	1 (0.5)	1 (0.3)	1 (0.3)
Sponsor decision	0	0	0	1 (0.3)
Other	0	1 (0.5)	0	1 (0.3)

[Source: Reviewer's Analysis]

Reviewer's Comment: Patient disposition was described in terms of the safety population in the CSR. To follow the clinical review, disposition is presented here using the ITT and PITT populations. Note that the safety population excludes subjects who were never treated. Additionally, one subject randomized to the everolimus arm who received only cabozantinib as a study treatment was evaluated in the cabozantinib arm in the safety population

3.2.3.2 Demographic and Baseline Characteristics

Demographics and select baseline characteristics for both the PITT and ITT populations are shown in Tables 3 and 4. All results appeared similar for the ITT and PITT populations.

Demographics and baseline characteristics, particularly the randomization stratification factors, were balanced between treatment arms in both populations.

Table 3: Demographics

	PITT Population		ITT Population	
	Cabozantinib N = 187	**Everolimus N = 188**	**Cabozantinib N = 330**	**Everolimus N = 328**
Age (years)				
Mean (standard deviation)	61.2 (9.29)	61.1 (10.39)	61.7 (9.51)	61.1 (10.50)
Median (range)	62.0 (36, 83)	61.0 (31, 84)	62.5 (32, 86)	62.0 (31, 84)
Age Category (years), n (%)				
< 65	118 (63)	116 (62)	196 (59)	198 (60)
≥ 65	69 (37)	72 (38)	134 (41)	130 (40)
Sex, n (%)				
Male	142 (76)	130 (69)	253 (77)	241 (73)
Female	45 (24)	57 (30)	77 (23)	86 (26)
Missing	0	1 (0.5)	0	1 (0.3)
Race[a], n (%)				
White	157 (84)	147 (78)	269 (82)	263 (80)
Asian	12 (6.4)	20 (11)	21 (6.4)	26 (7.9)
Black or African American	4 (2.1)	2 (1.1)	6 (1.8)	3 (0.9)
Other	10 (5.3)	6 (3.2)	19 (5.8)	13 (4.0)
Not reported	4 (2.1)	12 (6.4)	15 (4.5)	22 (6.7)
Missing	0	1 (0.5)	0	1 (0.3)
Ethnicity, n (%)				
Hispanic or Latino	12 (6.4)	10 (5.3)	19 (5.8)	18 (5.5)
Not Hispanic or Latino	160 (86)	160 (85)	278 (84)	273 (83)
Not reported	15 (8.0)	17 (9.0)	33 (10)	36 (11)
Missing	0	1 (0.5)	0	1 (0.3)
Geographic region, n (%)				
Europe	83 (44)	84 (45)	167 (51)	153 (47)
North America	76 (41)	64 (34)	118 (36)	122 (37)
Asia Pacific	25 (13)	36 (19)	39 (12)	47 (14)
Latin America	3 (1.6)	4 (2.1)	6 (1.8)	6 (1.8)

[a] Subjects could report more than one race.
[Source: CSR Table 15]

Table 4: Baseline Characteristics

	PITT Population		ITT Population	
	Cabozantinib N = 187	Everolimus N = 188	Cabozantinib N = 330	Everolimus N = 328
Randomization Stratification Factors per CRF, n (%)				
Prior VEGFR-TKI = 1	137 (73)	136 (72)	235 (71)	229 (70)
Prior VEGFR-TKI ≥ 2	50 (27)	52 (28)	95 (29)	99 (30)
MSKCC risk factors = 0 (favorable)	80 (43)	83 (44)	150 (45)	150 (46)
MSKCC risk factors = 1 (intermediate)	80 (43)	75 (40)	139 (42)	135 (41)
MSKCC risk factors = 2 or 3 (poor)	27 (14)	30 (16)	41 (12)	43 (13)
Karnofsky performance status (KPS)[a], n (%)				
70	15 (8.0)	16 (8.5)	29 (8.8)	22 (6.7)
80	43 (23)	56 (30)	75 (23)	90 (27)
90	72 (39)	81 (43)	127 (38)	142 (43)
100	57 (30)	35 (19)	99 (30)	74 (23)
Diagnosis of RCC with a clear cell component by histology or cytology, n (%)	187 (100)	187 (99)[b]	330 (100)	327 (100)[b]
Time since initial histological/cytological diagnosis to randomization				
< 1 year, n (%)	34 (18)	44 (23)	59 (18)	76 (23)
≥ 1 year, n (%)	153 (82)	143 (76)	271 (82)	251 (77)
Median (range) (years)	2.6 (0, 30)	2.4 (0, 33)	2.8 (0, 30)	2.5 (0, 33)
Current disease stage, n (%)				
Stage IV	153 (82)	166 (88)	272 (82)	287 (88)
Stage III	20 (11)	13 (6.9)	34 (10)	24 (7.3)
Unknown	14 (7.5)	8 (4.3)	24 (7.3)	16 (4.9)
Extent of baseline disease by IRC, n (%)				
Visceral	139 (74)	142 (76)	241 (73)	245 (75)
Lung	115 (61)	126 (67)	204 (62)	212 (65)
Liver	52 (28)	58 (31)	88 (27)	103 (31)
Brain	2 (1.1)	1 (0.5)	2 (0.6)	1 (0.3)
Lymph Node	124 (66)	110 (59)	206 (62)	199 (61)
Kidney	46 (25)	36 (19)	70 (21)	66 (20)
Bone (CT or MRI)	39 (21)	32 (17)	77 (23)	65 (20)
Other	16 (8.6)	10 (5.3)	23 (7.0)	21 (6.4)
Number of involved organs by IRC, n (%)				
1	31 (17)	31 (16)	59 (18)	56 (17)
2	57 (30)	48 (26)	101 (31)	77 (23)
≥ 3	98 (52)	105 (56)	168 (51)	190 (58)
Missing	1 (0.5)	4 (2.1)	2 (0.6)	5 (1.5)
Sum of Lesion Diameters (SoD) (mm)				
N	187	187	330	327
Mean (SD)	81.3 (56.81)	87.8 (55.65)	77.4 (54.73)	81.2 (53.74)
Median (range)	70.0 (0, 291)	77.0 (0, 231)	65.2 (0, 291)	65.0 (0, 258)
Type of prior VEGFR-TKI, n (%)				
Sunitinib	114 (61)	113 (60)	210 (64)	205 (63)
Pazopanib	87 (47)	78 (41)	144 (44)	136 (41)
Axitinib	28 (15)	28 (15)	52 (16)	55 (17)
Sorafenib	11 (5.9)	19 (10)	21 (6.4)	31 (9.5)

[a] KPS: 100 (normal activity), 90 (normal activity, minor signs and symptoms), 80 (normal activity with effort, some signs and symptoms), 70 (unable to carry on normal activity or to work, cares for self).

[b] One subject had RCC of unclassified (undifferentiated) histology and is not included in the numerator. Another subject in the everolimus arm had a histology of clear cell that could not be verified by the because of limited tissue, but a clear cell histology was favored; this subject is included in the numerator
[Source: CSR Tables 16]

Reviewer's Comment: The primary PFS analysis by the Sponsor was stratified by data collected on the eCRFs. A concordance analysis of eCRF- and IVRS-based stratification values revealed some discrepancies, more so for the MSKCC risk factors than the prior VEGFR-TKIs, as summarized in Tables 5 and 6. Concordance is summarized in Table 7. Discrepancies were consistent between the ITT and PITT populations. To follow the intent-to-treat principle, IVRS data should be used in the stratified analysis. See Section 3.2.4.1 for a sensitivity analysis of PFS using IVRS-based stratification data.

Table 5: Number of Prior VEGFR-TKIs Concordance between eCRF and IVRS

	PITT Population				ITT Population			
	Cabozantinib N =187		Everolimus N = 188		Cabozantinib N = 330		Everolimus N = 328	
	IVRS, n (%)							
eCRF, n (%)	1	≥ 2	1	≥ 2	1	≥ 2	1	≥ 2
1	134 (72)	3 (1.6)	134 (71)	2 (1.1)	230 (70)	5 (1.5)	227 (69)	2 (0.6)
≥ 2	2 (1.1)	48 (26)	2 (1.1)	50 (27)	3 (0.9)	92 (28)	4 (1.2)	95 (29)

[Source: Reviewer's Analysis]

Table 6: Number of MSKCC Risk Factors Concordance between eCRF and IVRS

	PITT Population						ITT Population					
	Cabozantinib N =187			Everolimus N = 188			Cabozantinib N = 330			Everolimus N = 328		
	IVRS, n (%)											
eCRF, n (%)	0	1	2 or 3	0	1	2 or 3	0	1	2 or 3	0	1	2 or 3
0 (favorable)	74 (40)	6 (3.2)	0	75 (40)	8 (4.3)	0	139 (42)	10 (3.0)	1 (0.3)	137 (42)	13 (4.0)	0
1 (intermediate)	7 (3.7)	69 (37)	4 (2.1)	5 (2.7)	68 (36)	2 (1.1)	10 (3.0)	124 (38)	5 (1.5)	11 (3.4)	119 (36)	5 (1.5)
2 or 3 (poor)	1 (0.5)	2 (1.1)	24 (13)	2 (1.1)	3 (1.6)	25 (13)	1 (0.3)	3 (0.9)	37 (11)	2 (0.6)	4 (1.2)	37 (11)

[Source: Reviewer's Analysis]

Table 7: Concordance between eCRF and IVRS stratification data

	PITT Population		ITT Population	
	Cabozantinib N =187	Everolimus N = 188	Cabozantinib N = 330	Everolimus N = 328
Number of Prior VEGFR-TKIs				
Concordance between eCRF and IVRS	182 (97%)	184 (98%)	322 (98%)	322 (98%)
Discordance between eCRF and IVRS	5 (3%)	4 (2%)	8 (2%)	6 (2%)
Number of risk factors per MSKCC prognostic criteria for previously treated patients with RCC				
Concordance between eCRF and IVRS	167 (89%)	168 (89%)	300 (91%)	293 (89%)
Discordance between eCRF and IVRS	20 (11%)	20 (11%)	30 (9%)	35 (11%)

[Source: Reviewer's Analysis]

3.2.3.3 Post-Study Treatment Anti-Cancer Therapy

Any non-protocol anticancer therapy (NPACT) was not to be initiated until after study treatment had been discontinued. Local anticancer treatment including palliative radiation, ablation, embolization, or surgery with impact on tumor lesions was not permitted (unless Sponsor-approved if unavoidable) until radiographic tumor assessments had been discontinued per protocol defined criteria.

Table 8 summarizes concomitant and subsequent chemotherapy. Generally, more subjects received NPACT in the everolimus arm than in the cabozantinib arm. More subjects in the everolimus arm received kinase inhibitors but more subjects in the cabozantinib arm received commercial everolimus.

Table 8: Concomitant and Subsequent Chemotherapy

	PITT Population		ITT Population	
	Cabozantinib	**Everolimus**	**Cabozantinib**	**Everolimus**
	N =187	**N = 188**	**N = 330**	**N = 328**
Everolimus	52 (28%)	7 (4%)	75 (23%)	13 (4%)
Kinase inhibitors	36 (20%)	84 (45%)	54 (16%)	133 (41%)
Monoclonal antibodies	5 (3%)	4 (2%)	6 (2%)	8 (2%)
Pyrimidine analogues	4 (2%)	2 (1%)	5 (2%)	3 (1%)
Interferons	3 (2%)	3 (2%)	3 (1%)	6 (2%)
Nitrogen mustard analogues	1 (1%)	1 (1%)	1 (<1%)	1 (<1%)
Ixabepalone	1 (1%)	0	1 (<1%)	0
Platinum compounds	0	1 (1%)	1 (<1%)	1 (1%)
Vinblastine	1 (1%)	1 (1%)	1 (<1%)	1 (<1%)
Interleukins	0	3 (2%)	0	4 (1%)

[Source: FDA Clinical Review]

Reviewer's Comment: The study did not allow for crossover, but note that there were five subjects on the everolimus arm in the ITT population that received commercial cabozantinib as NPACT (data not presented here).

3.2.3.4 Protocol Deviations

Protocol deviations in this study were generally well balanced between arms and the clinical review team believes they are unlikely to have had a substantial effect on efficacy. Refer to the clinical review for more details

3.2.4 Results and Conclusions

3.2.4.1 Primary Efficacy Endpoint

The pre-specified primary PFS analysis per IRC was based on the first 375 subjects randomized (PITT population) with a data cutoff date of 22 May 2015 at which point a total of 247 events were reported. The minimum time of follow-up was 10.7 months. Results are presented in Table 9. There was a statistically significant improvement in PFS for subjects in the cabozantinib arm compared with the everolimus arm, with a 3.6 month difference in median PFS (7.4 vs. 3.8 months) and the hazard ratio adjusted for stratification factors was 0.59 (95% CI: 0.46, 0.76;

stratified log-rank p-value < 0.0001) per eCRF and 0.58 (95% CI: 0.45, 0.74; stratified log-rank p-value < 0.0001) per IVRS.

As a sensitivity analysis, the applicant repeated the PFS per IRC analysis using the full ITT population which had a total of 394 events by the 22 May 2015 cutoff date with a minimum time of follow up of 5.9 months. Results are also shown in Table 9. The cabozantinib arm had a statistically significant 3.5 month median PFS improvement over the everolimus arm (7.4 vs. 3.9 months) with a stratified hazard ratio of 0.52 (95% CI: 0.43, 0.64; stratified log-rank p-value < 0.0001) per eCRF and 0.51 (95% CI: 0.41, 0.62; stratified log-rank p-value < 0.0001) per IVRS. Results from this analysis are consistent with the PITT population with a similar median PFS improvement (3.6 months in PITT vs. 3.5 months in ITT). The minimum follow up in the ITT was understandably shorter (5.9 months vs. 10.7 months in PITT). The magnitude of benefit represented by the stratified hazard ratio was larger in ITT. Kaplan-Meier curves for both populations are shown in Figure 2.

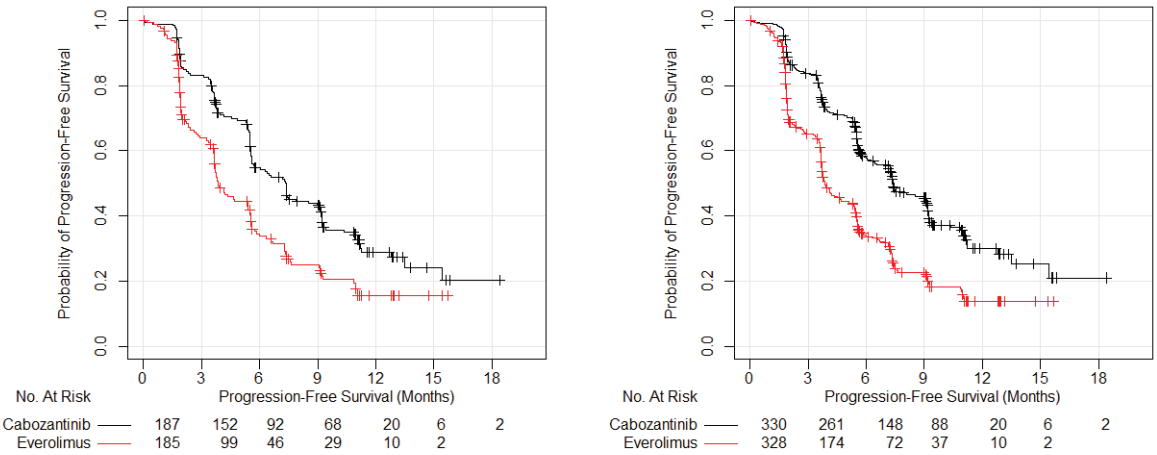


Figure 2: Kaplan-Meier Plot of Progression-Free Survival through the 22 May 2015 Cutoff Date per IRC in the PITT Population (left) and ITT Population (right)
[Source: CSR Figure 3]

Table 9: Progression-Free Survival Analysis per IRC

	PITT Population		ITT Population	
	Cabozantinib N = 187	Everolimus N = 188	Cabozantinib N = 330	Everolimus N = 328
Number (%) of subjects				
Censored	66 (35)	62 (33)	150 (45)	114 (35)
2 or more missed ATA prior to event	1 (0.5)	5 (2.7)	1 (0.3)	6 (1.8)
Anticancer therapy	24 (13)	31 (16)	35 (11)	49 (15)
No event by last ATA	39 (21)	23 (12)	104 (32)	50 (15)
No post-baseline ATA[a]	0	3 (1.6)	2 (0.6)	7 (12.1)
Surgery	2 (1.1)	0	8 (2.4)	2 (0.6)
Event	121 (65)	126 (67)	180 (55)	214 (65)
Death	8 (4.3)	13 (6.9)	16 (4.8)	18 (5.5)
Documented progression	113 (60)	113 (60)	164 (50)	196 (60)
Duration of progression-free survival (months)				
Median (95% CI)	7.4 (5.6, 9.1)	3.8 (3.7, 5.4)	7.4 (6.6, 9.1)	3.9 (3.7, 5.1)
25th percentile, 75th percentile[b]	3.7, 13.5	1.9, 9.1	3.7, 15.4	1.9, 7.4
Range	0.03, 18.4	0.03, 15.7	0.03, 18.4	0.03, 15.7
p-value (stratified log-rank per eCRF)	< 0.0001		< 0.0001	
Hazard ratio (95% CI; stratified per eCRF)	0.59 (0.46, 0.76)		0.52 (0.43, 0.64)	
p-value (stratified log-rank per IVRS)	< 0.0001		< 0.0001	
Hazard ratio (95% CI; stratified per IVRS)	0.58 (0.45, 0.74)		0.51 (0.41, 0.62)	
p-value (unstratified log-rank test)	< 0.0001		< 0.0001	
Hazard ratio (95% CI; unstratified)	0.59 (0.46, 0.76)		0.52 (0.42, 0.63)	

[a] Two of the subjects on the everolimus arm with no post-baseline ATA did not receive study treatment.
[b] Median and percentiles are based on Kaplan-Meier survival estimates.
[Source: CSR Tables 21, 23]

Reviewer's Comment: The study accrued faster than expected and the longer follow-up in the PITT population did not have a large impact on results.

Additional Sensitivity Analyses

The applicant performed additional sensitivity analyses using alternative definitions of progression events, as listed below.

1. Uniform dates: Used the scheduled tumor assessment date (or the next scheduled tumor assessment date if between assessments) rather than the date progression was recorded by the IRC as the date of radiographic progression.
2. Investigator claims: Events defined as the earliest of death, radiographic progression as assessed by the investigator, clinical deterioration, initiation of subsequent anticancer therapy, and surgery that impacted tumor lesions.
3. Investigator-documented Radiographic PD: Events defined as earlier of death or radiographic progression as determined by the investigator. Clinical deterioration was not considered a progression event.

The applicant also performed sensitivity analyses using alternative censoring schemes to explore the effect of potentially informative censoring resulting from (a) discontinuation of radiographic

assessments prior to progression/receipt of subsequent treatments or (b) progression by investigator prior to receipt of subsequent treatments or (c) receipt of subsequent treatments prior to progression, as listed below.

1. Censoring Scheme 1: Subjects with criteria (a) and (b) were classified as events in the cabozantinib arm only.
2. Censoring Scheme 2: Subjects in the cabozantinib arm meeting critera (a), (b), and (c) and those in the everolimus arm satisfying criteria (c) were classified as events.
3. Censoring Scheme 3: Subjects in the cabozantinib arm meeting criteria (a) and (b) and those in the everolimus arm satisfying criteria (b) were classified as events.
4. Censoring Scheme 4: All subjects meeting criteria (a), (b), and (c) were counted as events in the cabozantinib arm and remained censored in the everolimus arm.

Results from all of these analyses in both populations are summarized in Tables 10 and 11 and all show a longer PFS in the cabozantinib arm. Note that the primary analysis of PFS based on IRC and PFS based on Investigator-documented Radiographic PD were consistent. Censoring scheme 4 was considered the worst-case scenario analysis and the results reflect that.

Table 10: Sensitivity Analyses of PFS in the PITT Population

Sensitivity Analysis	**Cabozantinib**	**Everolimus**	**Stratified HR[a]**
	Median PFS (months)		(95% CI)
Uniform dates	7.4	3.9	0.59 (0.45, 0.76)
Investigator claims	7.3	4.0	0.59 (0.47, 0.74)
Investigator-documented Radiographic PD	7.4	5.3	0.61 (0.48, 0.77)
Censoring Scheme 1	5.7	3.8	0.73 (0.57, 0.92)
Censoring Scheme 2	5.6	3.7	0.70 (0.56, 0.89)
Censoring Scheme 3	5.7	3.7	0.58 (0.46, 0.73)
Censoring Scheme 4	5.6	3.8	0.75 (0.59, 0.95)

[a] Stratified by eCRF values
[Source: CSR Tables 24, 25]

Table 11: Sensitivity Analyses of PFS in the ITT Population

Sensitivity Analysis	**Cabozantinib**	**Everolimus**	**Stratified HR[a]**
	Median PFS (months)		(95% CI)
Uniform dates	7.4	3.9	0.52 (0.42, 0.64)
Investigator claims	7.3	3.9	0.53 (0.45, 0.64)
Investigator-documented Radiographic PD	7.4	5.1	0.54 (0.45, 0.66)
Censoring Scheme 1	6.0	3.9	0.66 (0.54, 0.79)
Censoring Scheme 2	5.8	3.7	0.65 (0.54, 0.78)
Censoring Scheme 3	6.0	3.7	0.53 (0.45, 0.64)
Censoring Scheme 4	5.8	3.9	0.69 (0.57, 0.84)

[a] Stratified by eCRF values
[Source: Reviewer's Analysis]

<u>Reviewer's Comment:</u> Data was available to repeat these sensitivity analyses using the ITT population but results were not explicitly shown in the CSR. Results from our analysis of the ITT population are shown in Table 11 and appear to be consistent with the PITT population.

FDA Sensitivity Analyses

The applicant had concerns that a primary PFS analysis using the first 259 PFS events amongst all randomized subjects (the ITT population) would potentially be biased due to the influence of subjects with early disease progression. Another concern was that the 259th event could have occurred prior to the study being completely accrued. We determined that the 259th PFS event occurred on 20 November 2014, which was just prior to when the last subject was randomized on 24 November 2014. Since these two dates were not far apart, we performed a post-hoc analysis of PFS using the ITT population with a cutoff date of 24 November 2014, at which point there were 263 PFS events. Some patients had very short follow-up time as the cutoff date was set to the date the last subject was randomized. There was a 2.3 month improvement in median PFS (6.0 vs. 3.7 months) for the cabozantinib arm over the everolimus arm. The stratified hazard ratio (per IVRS) was 0.49 (95% CI: 0.38, 0.63; stratified log-rank nominal p-value < 0.0001) and the unstratified hazard ratio was 0.50 (95% CI: 0.39, 0.64; log-rank nominal p-value < 0.0001).

Of the 263 subjects who had progressed by the 24 November 2014 cutoff date, 207 were counted as having had disease progression in the PITT population as well. The minimum time of follow up for these 207 subjects was 4.9 months. The 56 subjects who did not overlap with the PITT population had a minimum time of follow up of 1.6 months.

Compared with the pre-specified primary PFS analysis, the median PFS improvement was smaller but the magnitude of treatment effect was larger. The Kaplan-Meier Plot in Figure 3 reflects the influence of more subjects with early disease progression. The hazard ratio confidence intervals in this analysis (Table 12) are wider when compared with the ITT PFS analysis using a 22 May 2015 cutoff date (Table 9).

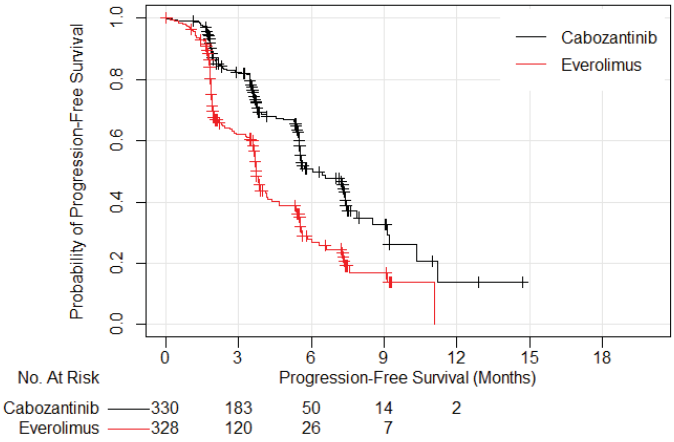


Figure 3: Kaplan-Meier Plot of Progression-Free Survival through the 24 November 2014 Cutoff Date per IRC in the ITT Population (263 PFS Events)

[Source: Reviewer's Analysis]

Table 12: Exploratory Progression-Free Survival Analysis per IRC in the ITT population (24 November 2014 Cutoff Date when 263 PFS Events Occurred)

	Cabozantinib N = 330	Everolimus N = 328
Number (%) of subjects		
Censored	218 (66)	177 (54)
2 or more missed ATA prior to event	1 (0.3)	6 (1.8)
Anticancer therapy	18 (5.5)	37 (11)
No event by last ATA	192 (58)	126 (38)
No post-baseline ATA	2 (0.6)	7 (2.1)
Surgery	5 (1.5)	1 (0.3)
Event	112 (34)	151 (46)
Death	13 (3.9)	15 (4.6)
Documented progression	99 (30)	136 (41)
Duration of progression free survival (months)		
Median (95% CI)	3.7 (3.6, 4.1)	6.0 (5.6, 7.4)
25th percentile, 75th percentile[a]	1.9, 6.6	3.6, 10.3
Range	0.03, 14.7	0.03, 11.1
Nominal p-value (stratified log-rank per eCRF)	< 0.0001	
Hazard ratio (95% CI; stratified per eCRF)	0.49 (0.38, 0.63)	
Nominal p-value (stratified log-rank per IVRS)	< 0.0001	
Hazard ratio (95% CI; stratified per IVRS)	0.49 (0.38, 0.63)	
Nominal p-value (unstratified log-rank test)	< 0.0001	
Hazard ratio (95% CI; unstratified)	0.50 (0.39, 0.64)	

[a] Median and percentiles are based on Kaplan-Meier survival estimates.
[Source: Reviewer's Analysis]

FDA also performed the following additional sensitivity analyses:

1. Using the Safety population which excludes subjects on the everolimus arm who did not receive treatment.
2. Using stratification data from the IVRS rather than the eCRF in the PFS per IRC analysis.
3. Treating IRC determined radiographic PD as an event regardless of censoring (i.e. even after two or more missing tumor assessments).
4. Combining IRC and investigator assessments: If both IRC and investigator agreed on a PFS event or censoring, the time to event or censoring time was set as the earlier of the two times. If only one called a PFS event, then that time to event was used.

Results are shown in Table 13 and support the robustness of the primary PFS findings.

Table 13: FDA Additional Sensitivity Analyses

Sensitivity Analysis	Cabozantinib	Everolimus	Stratified HR[a]
	Median PFS (months)		(95% CI)
Safety Population			
PITT Population	7.4	3.8	0.59 (0.46, 0.76)
ITT Population	7.4	3.9	0.52 (0.43, 0.64)
Stratified by IVRS values			
PITT Population	7.4	3.8	0.58 (0.45, 0.74)
ITT Population	7.4	3.9	0.51 (0.41, 0.62)
All IRC radiographic PDs as events			
PITT Population	6.9	3.8	0.59 (0.46, 0.75)
ITT Population	7.4	3.9	0.53 (0.44, 0.64)
Combination of IRC and Investigator			
PITT Population	5.6	3.7	0.59 (0.47, 0.74)
ITT Population	5.8	3.7	0.53 (0.44, 0.63)

[a] Stratified by eCRF values unless otherwise specified
[Source: Reviewer's Analysis]

FDA Analysis of Concordance between IRC and Investigator Assessment of Radiographic Progressive Disease

The applicant's sensitivity analyses showed that results for PFS based on IRC-determined and investigator-determined radiographic PD were similar. In this section we assess the concordance between IRC and investigator to more closely investigate the extent of investigator bias present in this open-label study.

The applicant's concordance analysis reported more progression events than there were in the respective PFS analyses because censoring was not taken into account. Results from that analysis are summarized in Appendix B. FDA performed a concordance analysis using events (death or PD) and censored observations as reported in the PFS analyses. Results are shown in Tables 14 and 15.

In the PITT population, the IRC and investigator agreed 76% of the time for the cabozantinib arm and 72% of the time for the everolimus arm. In the ITT population, they agreed 77% of the time for the cabozantinib arm and 74% of the time for the everolimus arm.

In the PITT population, when both IRC and investigator agreed on PD, they agreed on the dates of PD 46% of the time for the cabozantinib arm and 59% of the time for the everolimus arm. In the ITT population, they agreed on the dates of PD 49% of the time for the cabozantinib arm and 60% of the time for the everolimus arm.

Table 14: FDA Analysis of Concordance between IRC and Investigator Read in Progressive Disease Status for Tumor Assessment

		Cabozantinib N = 187 (PITT), N = 330 (ITT)			Everolimus N = 188 (PITT), N = 328 (ITT)		
		IRC Read, n (%)					
	Investigator Read, n (%)	Censored	Death	PD	Censored	Death	PD
PITT Population	Censored	40 (21)	0	14 (7.5)	32 (17)	0	15 (8.0)
	Death	0	7 (3.7)	3 (1.6)	0	10 (5.3)	4 (2.1)
	PD	26 (14)	1 (0.5)	96 (51)	30 (16)	3 (1.6)	94 (50)
ITT Population	Censored	107 (32)	0	27 (8.2)	68 (21)	0	27 (8.2)
	Death	0	14 (4.2)	3 (0.9)	0	12 (3.7)	7 (2.1)
	PD	43 (13)	2 (0.6)	134 (41)	46 (14)	6 (1.8)	162 (49)

[Source: Reviewer's Analysis]

Table 15: FDA Analysis of Concordance between IRC and Investigator Read in Date of PD for Tumor Assessments Among Subjects Who Progressed

	Concordance	**Cabozantinib N = 96**	**Everolimus N = 94**	**Total N = 190**
PITT Population	Yes	44 (46)	55 (59)	99 (52)
	No	52 (54)	39 (41)	91 (48)
	Investigator before IRC	14	6	20
	Investigator after IRC	38	33	71
	Concordance	**Cabozantinib N = 134**	**Everolimus N = 162**	**Total N =296**
ITT Population	Yes	66 (49)	98 (60)	164 (55)
	No	68 (51)	64 (40)	132 (45)
	Investigator before IRC	20	13	33
	Investigator after IRC	48	51	99

[Source: Reviewer's Analysis]

Overall, there was a slightly higher rate of concordance in PD status for the cabozantinib arm compared to the everolimus arm. However, when considering date of PD, there was more discordance in the cabozantinib arm. To assess any possible investigator bias, we apply the PhRMA method of differential discordance (Amit et al. 2011, Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis, *European Journal of Cancer*, 47, 1772-8*)*. The method takes into account the early discrepancy rate (EDR), defined as the rate of investigator determined progression events that occurred earlier than IRC as a proportion of the number of investigator progression events, and the late discrepancy rate (LDR), defined as the rate of investigator determined progression events that occurred later than IRC as a proportion of the number of discordances. Results for this study are shown in Table 16.

Table 16: Differential Discordance Analysis

		Cabozantinib	**Everolimus**	**Difference**
PITT Population	EDR	33.3%	30.7%	2.6%
	LDR	57.3%	57.1%	0.2%
ITT Population	EDR	36.3%	30.4%	5.9%
	LDR	54.5%	56.7%	-2.2%

[Source: Reviewer's Analysis]

Typically a negative differential discordance for EDR indicates that the investigator was more likely to call PD before IRC in the control arm and a positive differential discordance for LDR indicates that the investigator was more likely to call PD after IRC in the experimental arm. Evidence of either would be grounds for concluding that the investigator was biased towards the cabozantinib arm. In this case differential discordance for EDR was positive in both populations while differential discordance for LDR was slightly positive in the PITT population and negative in the ITT population. Generally all these differences were small and suggest that, in retrospect, investigator bias did not appear to be a big issue even though the primary analysis was per IRC.

3.2.4.2 Key Secondary Endpoints

Interim Analysis of Overall Survival (ITT Population)

The applicant conducted the pre-specified interim analysis of OS on the ITT population (as of the 22 May 2015 database cutoff) at the time of the primary analysis of PFS. There were 202 total deaths by the cutoff date, representing 49% (202/408) of the total deaths required for the pre-specified primary analysis of OS. The minimum time of follow-up was 5.9 months. A longer OS for subjects in the cabozantinib arm compared to the everolimus arm was observed, with a stratified hazard ratio of 0.68 (95% CI: 0.51, 0.90; stratified log-rank p-value = 0.006) per eCRF and 0.68 (95% CI: 0.52, 0.90; stratified log-rank p-value = 0.007) per IVRS. The p-value of ≤ 0.0019 (49% information fraction) required to achieve statistical significance at the time of the interim analysis was not reached.

A second unplanned interim analysis of OS was conducted for the ITT population with a data cutoff date of 31 December 2015. There were 320 deaths by this cutoff date, representing 78% of the 408 total deaths required for the pre-specified primary analysis of OS. The minimum time of follow-up was 13 months. There was a 4.9 month improvement in median OS (21.4 vs. 16.5 months) for subjects in the cabozantinib arm compared with the everolimus arm. The hazard ratio adjusted for stratification factors was 0.67 (95% CI: 0.53, 0.83; stratified log-rank p-value = 0.0003) per eCRF and 0.66 (95% CI: 0.53, 0.83; stratified log-rank p-value = 0.0003) per IVRS. The p-value of ≤ 0.0163 (78% information fraction) required to achieve statistical significance at the time of this unplanned second interim analysis was reached.

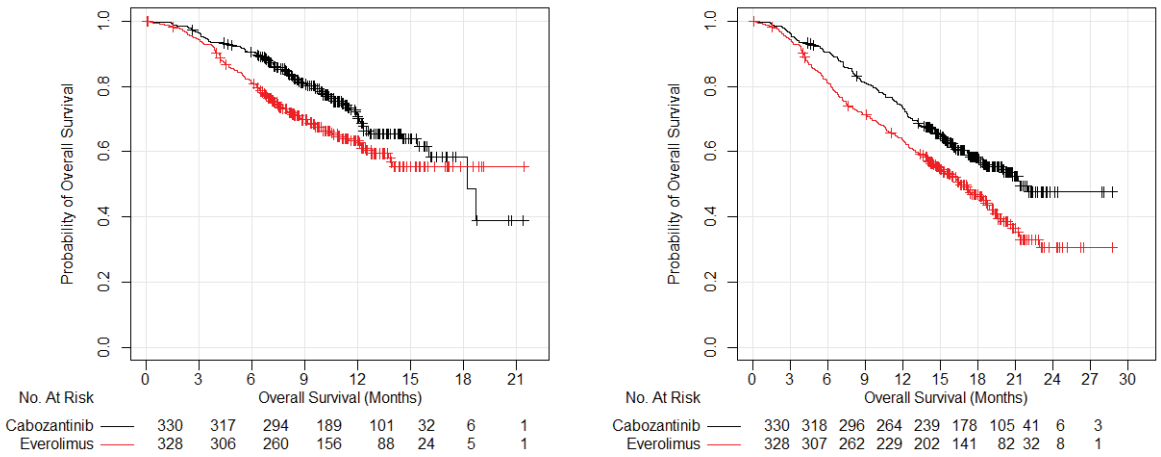


Figure 4: Kaplan-Meier Plot of Interim Overall Survival in the ITT Population through the 22 May 2015 cutoff (left) and the 31 December 2015 cutoff (right)
[Source: CSR Figure 5]

Table 17: Interim Analyses of Overall Survival in the ITT Population

	1st Interim (22 May 2015 cutoff)		Unplanned 2nd Interim (31 December 2015 cutoff)	
	Cabozantinib N = 330	Everolimus N = 328	Cabozantinib N = 330	Everolimus N = 328
Number (%) of subjects				
Censored	241 (73)	215 (66)	190 (58)	148 (45)
Death	89 (27)	113 (34)	140 (42)	180 (55)
Duration of overall survival (months)				
Median (95% CI)	18.2 (16.1, NE)	NE (13.9, NE)	21.4 (18.7, NE)	16.5 (14.7, 18.8)
25th percentile, 75th percentile[a]	11.1, NE	7.5, NE	11.5, NE	7.5, NE
Range	0.26, 21.4	0.03, 21.5	0.26, 28.7	0.07, 28.8
p-value (stratified log-rank per eCRF)	0.006		0.0003	
Hazard ratio (95% CI; stratified per eCRF)	0.68 (0.51, 0.90)		0.67 (0.53, 0.83)	
p-value (stratified log-rank per IVRS)	0.007		0.0003	
Hazard ratio (95% CI; stratified per IVRS)	0.68 (0.52, 0.90)		0.66 (0.53, 0.83)	
p-value (unstratified log-rank test)	0.010		0.0004	
Hazard ratio (95% CI; unstratified)	0.69 (0.53, 0.92)		0.67 (0.54, 0.84)	

[a] Median and percentiles are based on Kaplan-Meier survival estimates.
[Source: CSR Table 28, CSR Addendum Table 2, Reviewer's Analysis]

Reviewer's Comments:

1. *Stratified hazard ratios calculated per eCRF and per IVRS were consistent and both p-values reached the p-value required for statistical significance.*
2. *Although not a pre-specified analysis, the applicant also looked at OS in the PITT population and results were consistent with the ITT population.*

Objective Response Rate per IRC

The applicant's primary analysis of ORR was conducted in the ITT population at the time of primary analysis of PFS using the same data cutoff date. Tumor assessments that occurred after the individual subject PFS-censoring dates were excluded. Results in Table 18 showed that there was a statistically significant positive difference in confirmed ORR between cabozantinib and everolimus (17% vs. 3% confirmed partial responses; unstratified p-value < 0.001). Median time to objective response was 1.91 months (range 1.6, 11.0) in the cabozantinib arm and 2.14 months (range 1.9, 9.2) in the everolimus arm. The Kaplan-Meier estimate of median duration of response (DOR) was not estimable (95% CI: 7.2, NE) for the cabozantinib arm and 7.4 months (95% CI: 1.9, NE) for the everolimus arm.

There were 16 subjects in the cabozantinib arm and 26 subjects in the everolimus arm that were missing or unable to evaluate (UE). The primary reason for this was that seven subjects randomized to receive everolimus withdrew consent before the first scheduled tumor assessment, three of whom never received study treatment. Reasons for being missing or UE were similar in each arm otherwise. The applicant performed post-hoc sensitivity analyses analyses assuming those seven subjects had the same proportion of objective response as was observed in the remaining subjects randomized to everolimus and assuming all seven subjects were responders. Both analyses resulted in a nominal p-value < 0.0001.

Table 18: Tumor Response per RECIST 1.1 (22 May 2015 Cutoff Date) per IRC in the ITT Population

Subjects in ITT Population	**Cabozantinib N = 330**	**Everolimus N = 328**
Subjects with any tumor reduction compared with baseline, n (%)	249 (75)	158 (48)
Best overall response, n (%)		
Confirmed complete response (CR)	0	0
Confirmed partial response (PR)	57 (17)	11 (3)
Stable disease (SD)[a]	216 (65)	203 (62)
Progressive disease	41 (12)	88 (27)
Unable to evaluate (UE)	2 (0.6)	2 (0.6)
Missing[b]	14 (4)	24 (7)
Objective response rate (ORR)[c]		
n (%)	57 (17)	11 (3)
95% confidence interval, %	(13, 22)	(2, 6)
Stratified CMH test p-value, per eCRF	< 0.0001	
Stratified CMH test p-value, per IVRS	< 0.0001	
Unstratified chi-squared test p-value	< 0.0001	
Duration of Response (months) (95% CI)	NE (7.2, NE)	7.4 (1.9, NE)

[a] Includes subjects for whom the overall response result is stable disease or non-CR/non-PD.
[b] No qualifying post-baseline assessment for overall response.
[c] ORR is defined as the proportion of subjects achieving an overall response of CR or PR confirmed by a subsequent scan at least 28 days later.
[Source CSR Table 33]

Objective Response Rate per Investigator

The applicant also considered tumor response per investigator for the ITT population, as summarized in Table 19. The ORR was 24% (95% CI: 19, 29) and 4% (95% CI: 2, 7) in the

cabozantinib and everolimus arms, respectively. Median time to objective response was 1.91 months (range 1.3, 9.8) in the cabozantinib arm and 3.50 months (range 1.8, 5.6) in the everolimus arm. This analysis was not pre-specified but results are consistent with the ORR per IRC results. The Kaplan-Meier estimate of median duration of response (DOR) was 7.4 months (95% CI: 5.7, NE) for the cabozantinib arm and 7.4 months (95% CI: 5.6, NE) for the everolimus arm.

Table 19: Tumor Response per RECIST 1.1 (22 May 2015 Cutoff Date) per Investigator in the ITT Population

Subjects in ITT Population	Cabozantinib N = 330	Everolimus N = 328
Best overall response, n (%)		
Confirmed complete response (CR)	0	0
Confirmed partial response (PR)	78 (24)	14 (4)
Stable disease (SD)[a]	209 (63)	205 (63)
Progressive disease	29 (9)	87 (27)
Unable to evaluate	3 (0.9)	5 (2)
Missing[b]	11 (3)	17 (5)
Objective response rate (ORR)[c]		
n (%)	78 (24)	14 (4)
95% confidence interval, %	(19, 29)	(2, 7)
Stratified CMH test nominal p-value, per eCRF	< 0.0001	
Stratified CMH test nominal p-value, per IVRS	< 0.0001	
Unstratified chi-squared test nominal p-value	< 0.0001	
Duration of Response (months) (95% CI)	7.4 (5.7, NE)	7.4 (5.6, NE)

[a] Includes subjects for whom the overall response result is stable disease or non-CR/non-PD.
[b] No qualifying post-baseline assessment for overall response.
[c] ORR is defined as the proportion of subjects achieving an overall response of CR or PR confirmed by a subsequent scan at least 28 days later.
[Source: CSR Table 34]

3.2.4.3 Summary of Results in Additional Exploratory Endpoints

Bone scan response (BSR) in the cabozantinib arm was 18% (95% CI: 11, 27) compared to 10% (95% CI: 4, 19) in the everolimus arm. Median duration of BSR was not estimable in either treatment arm. For skeletal-related events (SRE), 12% of subjects in the cabozantinib arm and 14% of subjects in the everolimus arm had an SRE post-randomization. Given an SRE prior to randomization, the incidence of post-randomization SREs was lower in the cabozantinib arm (16%) than the everolimus arm (34%). Regarding health care resource utilization, hospitalization rates (37% vs. 40% of subjects; 6.4 vs. 10.2 days per person-year), ICU visit rates (1.2% vs. 2.1% of subjects; 0.07 vs. 0.32 days per person-year), and surgeries per person-year (0.90 vs. 1.35) were lower in the cabozantinib arm than the everolimus arm. Median hospitalizations, ICU visits, or surgeries per subject, among those who experienced these events, were similar in each treatment group.

Patient Reported Outcome Results

Two health-related quality of life (HRQOL) instruments were used: FKSI-19 and EQ-5D-5L. Assessments were taken every 4 weeks through week 25 and then every 8 weeks thereafter. The FKSI-19 instrument is a 19-item self-reported questionnaire that assesses change in kidney-

cancer related symptoms by querying symptom severity and interference in activity and general health perceptions. Each symptom was assessed by the subject on a 5 point scale from 0 (not at all) to 4 (very much) and scores were converted so that higher scores indicate improvement. Total scores and scores in four disease-related symptoms (DRS) subscales (FKSI-DRS-Physical, FKSI-DRS-Emotional, FKSI-Treatment Side Effects, and FKSI-Function/Well-Being) were derived. The EQ-5D-5L instrument assesses change in mobility, self-care, usual activities, pain/discomfort, anxiety/depression, and global health. The first five functional and symptom dimensions were scored at levels indicating increasing severity from Level 1 (no problem) to Levels 2 through 5 (mild problem to extreme problem). The EQ-5D-5L was converted into a single index value normalized across the nine different countries where it has been validated. Subjects also completed a 20-cm vertical visual assessment scale (VAS) scored from 0 ("worst health you can imagine") to 100 ("best health you can imagine").

The number of FKSI questionnaires completed dropped to approximately 50% of the original number of subjects by Week 20 in the everolimus arm and Week 32 in the cabozantinib arm. For EQ-5D-5L, the index questionnaire completion rate dropped to approximately 50% of the original number of subjects by Week 16 in the everolimus arm and Week 32 in the cabozantinib arm. The VAS questionnaire completion rate dropped to approximately 50% of the original number of subjects by Week 20 in the everolimus arm and by Week 32 in the cabozantinib arm.

For FKSI-19, a repeated-measures analysis of the change from baseline in total score, the four subscales, and the individual symptoms showed some measures favored cabozantinib and some favored everolimus, reflective of their safety profiles. There were no clinically important treatment differences for the FKSI-total score and three of the four subscales. The last subscale was FKSI-Treatment Side Effects, and it showed a worse score for cabozantinib for diarrhea which was the most frequent AE in that arm. For EQ-5D-5L, a repeated-measures analysis of the change from baseline in VAS and index scores showed no clinically meaningful treatment differences between cabozantinib and everolimus.

Reviewer's Comment: Please note that, although there were no meaningful differences in PRO outcomes in this open-label study, this does not mean that cabozantinib had no decrement in patients' health-related quality of life compared to everolimus since the Applicant did not plan to test specific hypothesis related to PRO outcomes. PRO results are subject to many limitations in an open-label study due to record bias, loss to follow up, and missing data, among other factors.

3.3 Evaluation of Safety

Please refer to the clinical evaluations of this application for safety results and conclusions related to safety.

3.4 Benefit-Risk Assessment

Please refer to the clinical evaluations of this application for benefit-risk evaluation

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

Results for PFS per IRC in the PITT population by gender, race, age, and geographic region are summarized in Table 20. Corresponding results for the ITT population are shown in Table 21 and results for OS in the ITT population are shown in Table 22. The PFS and OS benefit in the cabozantinib arm held across these various subgroups except for PFS in females in the PITT population and a select few small subgroups including Latin America (PITT: n = 7, ITT: n = 12), Black or African American (PITT: n = 6, ITT: n = 9), and Race Not Reported (PITT: n = 16, ITT: n = 37). Note that the PFS benefit was seen in females when considering the full ITT population. No outlier subgroup was observed.

Table 20: Exploratory Demographic Subgroup Analyses for PFS per IRC in the PITT Population

	N	Cabozantinib		Everolimus		Unstratified HR
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Age						
< 65 years	234	75/118 (64)	7.33	82/116 (71)	3.75	0.55 (0.40, 0.75)
>= 65 years	141	46/69 (67)	7.39	44/72 (61)	4.70	0.64 (0.42, 0.98)
Gender						
F	102	35/45 (78)	5.59	37/57 (65)	5.55	1.03 (0.65, 1.63)
M	272	86/142 (61)	7.89	88/130 (68)	3.71	0.48 (0.36, 0.65)
Geographic Region						
Asia Pacific	61	19/25 (76)	7.43	25/36 (69)	3.61	0.60 (0.33, 1.10)
Europe	167	51/83 (61)	7.33	62/84 (74)	3.84	0.54 (0.37, 0.79)
Latin America	7	3/3 (100)	11.04	2/4 (50)	NE	1.29 (0.21, 7.92)
North America	140	48/76 (63)	7.16	37/64 (58)	4.17	0.60 (0.39, 0.93)
Race						
Asian	32	8/12 (67)	7.43	14/20 (70)	2.83	0.52 (0.22, 1.26)
Black or African American	6	3/4 (75)	3.73	1/2 (50)	NE	0.70 (0.06, 7.92)
White	304	100/157 (64)	7.36	96/147 (65)	4.14	0.60 (0.45, 0.79)
Other	16	7/10 (70)	6.98	5/6 (83)	2.04	0.16 (0.04, 0.68)
Not Reported	16	3/4 (75)	5.57	9/12 (75)	6.01	1.22 (0.31, 4.76)

[Source: CSR Table 42 and Reviewer's Analysis]

Table 21: Exploratory Demographic Subgroup Analyses for PFS per IRC in the ITT Population

	N	Cabozantinib		Everolimus		Unstratified HR
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Age						
< 65 years	394	109/196 (56)	7.36	133/198 (67)	3.75	0.53 (0.41, 0.68)
>= 65 years	264	71/134 (53)	9.17	81/130 (62)	3.91	0.50 (0.36, 0.69)

Table 21: Exploratory Demographic Subgroup Analyses for PFS per IRC in the ITT Population (continued)

	N	Cabozantinib		Everolimus		Unstratified HR
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Gender						
F	163	48/77 (62)	5.75	55/86 (64)	4.70	0.72 (0.49, 1.07)
M	494	132/253 (52)	7.89	158/241 (66)	3.81	0.46 (0.36, 0.58)
Geographic Region						
Asia Pacific	86	21/39 (54)	9.20	33/47 (70)	3.61	0.43 (0.25, 0.75)
Europe	320	92/167 (55)	7.33	105/153 (69)	3.91	0.54 (0.41, 0.72)
Latin America	12	4/6 (67)	11.04	2/6 (33)	NE	1.38 (0.25, 7.66)
North America	240	63/118 (53)	7.36	74/122 (61)	4.11	0.50 (0.35, 0.70)
Race						
Asian	47	10/21 (48)	9.43	18/26 (69)	2.63	0.36 (0.16, 0.78)
Black or African American	9	4/6 (67)	3.73	1/3 (33)	NE	2.34 (0.26, 21.39)
White	532	142/269 (53)	7.89	169/263 (64)	3.91	0.50 (0.40, 0.63)
Other	32	13/19 (68)	6.57	9/13 (69)	3.68	0.42 (0.18, 1.01)
Not Reported	37	11/15 (73)	5.78	16/22 (73)	5.45	0.98 (0.45, 2.13)

[Source: CSR Table 42 and Reviewer's Analysis]

Table 22: Exploratory Demographic Subgroup Analyses for OS in the ITT Population at Unplanned 2nd Interim Analysis

	N	Cabozantinib		Everolimus		Hazard Ratio
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Age						
< 65 years	394	86/196 (44)	21.39	107/198 (54)	17.12	0.72 (0.54, 0.95)
>= 65 years	264	54/134 (40)	21.26	73/130 (56)	16.26	0.62 (0.44, 0.88)
Gender						
F	163	37/77 (48)	18.10	53/86 (62)	16.26	0.74 (0.48, 1.12)
M	494	103/253 (41)	22.01	126/241 (52)	17.22	0.66 (0.51, 0.85)
Geographic Region						
Asia Pacific	86	16/39 (41)	NE	32/47 (68)	12.78	0.49 (0.27, 0.90)
Europe	320	71/167 (43)	22.01	88/153 (58)	16.36	0.67 (0.49, 0.91)
Latin America	12	2/6 (33)	NE	3/6 (50)	NE	0.49 (0.08, 2.97)
North America	240	51/118 (43)	21.26	57/122 (47)	19.58	0.79 (0.54, 1.16)
Race						
Asian	47	7/21 (33)	21.39	15/26 (58)	15.18	0.45 (0.18, 1.12)
Black or African American	9	5/6 (83)	13.70	2/3 (67)	18.79	1.25 (0.23, 6.92)
White	532	110/269 (41)	NE	143/263 (54)	16.36	0.65 (0.51, 0.83)
Other	32	10/19 (53)	17.25	8/13 (62)	13.93	0.64 (0.25, 1.62)
Not Reported	37	8/15 (53)	18.07	11/22 (50)	17.31	1.63 (0.62, 4.26)

[Source: CSR Addendum Table 7 and Reviewer's Analysis]

4.2 Other Special/Subgroup Populations

Efficacy results of PFS per IRC and OS were also assessed by the stratification factors: number of prior VEGFR-TKI agents and number of MSKCC risk factors. Other subgroups of interest included duration of prior VEGFR-TKI treatment, MET status, and sunitinib or pazopanib as the only prior VEGFR inhibitor. PFS results are shown in Table 23 (PITT population) and Table 24 (ITT population). OS results in the ITT population are shown in Table 25. The PFS and OS benefit in the cabozantinib arm held across these various subgroups. Note that the subgroups of sunitinib or pazopanib as the only prior VEGFR-TKI therapy were post-hoc and not pre-specified. No outlier subgroup was observed.

Table 23: Additional Exploratory Subgroup Analyses for PFS per IRC in the PITT Population

	N	Cabozantinib		Everolimus		Hazard Ratio
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Number of prior VEGFR-TKIs per eCRF						
1	273	87/137 (64)	7.36	95/136 (70)	3.75	0.56 (0.42, 0.75)
>= 2	102	34/50 (68)	6.05	31/52 (60)	5.55	0.67 (0.41, 1.10)
Treatment Duration on first VEGFR-TKI						
<= 6 months	116	37/54 (69)	5.59	44/62 (71)	3.65	0.56 (0.36, 0.88)
> 6 months	259	84/133 (63)	7.43	82/126 (65)	4.37	0.59 (0.44, 0.81)
MSKCC Risk Factors per eCRF						
0	163	51/80 (64)	7.39	56/83 (67)	4.67	0.54 (0.37, 0.79)
1	155	49/80 (61)	7.39	47/75 (63)	3.71	0.56 (0.37, 0.84)
2 or 3	57	21/27 (78)	4.14	23/30 (77)	2.30	0.84 (0.46, 1.53)
Tumor MET IHC status						
High	56	19/30 (63)	7.36	18/26 (69)	3.65	0.48 (0.25, 0.92)
Low	173	54/83 (65)	6.41	59/90 (66)	4.70	0.69 (0.47, 1.00)
Unknown	146	48/74 (65)	8.97	49/72 (68)	3.71	0.53 (0.35, 0.79)
Only Prior VEGFR-TKI						
sunitinib	153	45/76 (59)	9.13	58/77 (75)	3.71	0.41 (0.28, 0.61)
pazopanib	104	38/55 (69)	6.41	30/49 (61)	5.39	0.81 (0.50, 1.31)

[Source: CSR Table 42]

Table 24: Additional Exploratory Subgroup Analyses for PFS per IRC in the ITT Population

	N	Cabozantinib		Everolimus		Hazard Ratio
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Number of prior VEGFR-TKIs per eCRF						
1	464	131/235 (56)	7.39	155/229 (68)	3.84	0.52 (0.41, 0.66)
>= 2	194	49/95 (52)	7.39	59/99 (60)	4.04	0.51 (0.35, 0.74)
Treatment Duration on first VEGFR-TKI						
<= 6 months	190	56/88 (64)	5.59	70/102 (69)	3.71	0.62 (0.44, 0.89)
> 6 months	466	124/242 (51)	8.97	142/224 (63)	3.91	0.48 (0.38, 0.62)
MSKCC Risk Factors per eCRF						
0	300	79/150 (53)	7.49	92/150 (61)	5.13	0.51 (0.38, 0.69)
1	274	74/139 (53)	7.46	89/135 (66)	3.75	0.47 (0.35, 0.65)
2 or 3	84	27/41 (66)	5.42	33/43 (77)	3.48	0.70 (0.42, 1.16)
Tumor MET IHC status						
High	96	26/48 (54)	7.36	36/48 (75)	3.71	0.41 (0.24, 0.68)
Low	289	79/138 (57)	7.16	97/151 (64)	4.14	0.58 (0.43, 0.79)
Unknown	273	75/144 (52)	9.10	81/129 (63)	3.71	0.50 (0.36, 0.68)
Only Prior VEGFR-TKI						
sunitinib	267	74/135 (55)	9.10	97/132 (73)	3.71	0.43 (0.32, 0.59)
pazopanib	171	51/88 (58)	7.36	49/83 (59)	5.13	0.67 (0.45, 0.99)

[Source: CSR Table 42]

Table 25: Additional Exploratory Subgroup Analyses for OS in the ITT Population at Unplanned 2nd Interim Analysis

	N	Cabozantinib		Everolimus		Hazard Ratio
		#event/n (%)	Median	#event/n (%)	Median	(95% CI)
Number of prior VEGFR-TKIs per eCRF						
1	464	98/235 (42)	21.39	130/229 (57)	16.53	0.65 (0.50, 0.85)
>= 2	194	42/95 (44)	20.80	50/99 (51)	17.22	0.73 (0.48, 1.10)
Treatment Duration on first VEGFR-TKI						
<= 6 months	190	42/88 (48)	21.26	65/102 (64)	13.77	0.69 (0.47, 1.01)
> 6 months	466	98/242 (40)	22.01	114/224 (51)	18.40	0.69 (0.52, 0.90)
MSKCC Risk Factors per eCRF						
0	300	48/150 (32)	NE	66/150 (44)	19.25	0.66 (0.46, 0.96)
1	274	64/139 (46)	19.94	79/135 (59)	14.85	0.67 (0.48, 0.94)
2 or 3	84	28/41 (68)	10.45	35/43 (81)	6.47	0.65 (0.39, 1.07)

Table 25: Additional Exploratory Subgroup Analyses for OS in the ITT Population at Unplanned 2nd Interim Analysis (continued)

	N	Cabozantinib		Everolimus		Hazard Ratio
		#event/n (%)	Median	#event/n (%)		
Tumor MET IHC status						
High	101	20/51 (39)	22.01	27/50 (54)	15.18	0.55 (0.31, 0.99)
Low	312	63/150 (42)	20.80	87/162 (54)	18.37	0.72 (0.52, 1.00)
Unknown	245	57/129 (44)	21.26	66/116 (57)	14.98	0.67 (0.47, 0.95)
Only Prior VEGFR-TKI						
sunitinib	267	59/135 (44)	21.39	80/132 (61)	16.46	0.66 (0.47, 0.93)
pazopanib	171	34/88 (39)	22.01	42/83 (51)	17.48	0.66 (0.42, 1.04)

[Source: CSR Addendum Table 7]

Reviewer's Comment: All subgroup analyses presented are considered exploratory or hypothesis generating and no formal inference can be drawn.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

Consideration of PFS as the primary endpoint for efficacy towards drug approval is based on the magnitude of effect and the risk/benefit profile of the product. PFS per IRC was the primary endpoint in this pivotal study. The cabozantinib arm showed a 3.6 month improvement in median PFS with a stratified hazard ratio of 0.58 (95% CI: 0.45, 0.74) per IVRS in the PITT population over everolimus. At the time of the primary PFS analysis, the pre-specified interim analysis of OS showed longer survival in the cabozantinib arm with a stratified hazard ratio of 0.68 (95% CI: 0.52, 0.90) per IVRS, but the critical p-value necessary to determine significance at interim was not reached. A second unplanned interim analysis at 78% information resulted in a statistically significant 4.9 month improvement in median OS in the cabozantinib arm with a stratified hazard ratio of 0.66 (95% CI: 0.53, 0.83) per IVRS.

The applicant pre-specified the primary analysis to be conducted on the PITT population (consisting of the first 375 randomized subjects) to allow for longer PFS follow up among a smaller number of subjects and to avoid over-representing subjects whose disease progressed early. An FDA sensitivity analysis indicated that performing the primary analysis using the first 263 PFS events in the full ITT population would have resulted in a 2.3 month improvement in median PFS from the cabozantinib (stratified HR per IVRS = 0.49; 95% CI: 0.38, 0.63). The smaller PFS improvement may have been a result of the suspected bias towards patients whose disease progressed early.

The primary PFS analysis was also repeated using the full ITT population as a sensitivity analysis. Though this analysis was technically overpowered, since the 394 PFS events at data cutoff was more than the 259 necessary for an event-driven analysis, a similar statistically significant median PFS improvement (3.5 months) was seen with a stratified hazard ratio of 0.51 (95% CI: 0.41, 0.62) per IVRS.

A number of other sensitivity analyses were conducted by the applicant as well as the FDA and all supported the robustness of the primary PFS results.

5.2 Collective Evidence

There was a statistically significant improvement in PFS per IRC in the cabozantinib arm over the everolimus arm in both the PITT and ITT populations. The robustness of these results was supported by a number of sensitivity analyses conducted by the applicant as well as the FDA. Both interim analyses of overall survival showed longer survival in the cabozantinib arm, with the second unplanned analysis showing a statistically significant difference. There was also a statistically significant positive difference in confirmed ORR of cabozantinib over everolimus.

5.3 Conclusions and Recommendations

Based on the evidence from the METEOR study, cabozantinib appears to have a PFS, OS, and ORR benefit over everolimus. We recommend that the stratified hazard ratios for PFS and OS reported in the label should reflect the IVRS values to follow the ITT principle. The final decision on the benefit/risk profile in support of approval of cabozantinib in this setting is deferred to the clinical review team.

APPENDIX A: Schedule of Assessments

Both PFS and ORR were based on RECIST 1.1 per IRC. CT (or MRI) of chest/abdomen/pelvis (CAP) was performed in all subjects at screening, every 8 weeks (± 5 days) after randomization throughout the first 12 months on study, and then every 12 weeks afterwards (± 7 days). If MRI of the abdomen and pelvis was performed at screening, then a CT of the chest was as well. MRI (or CT) of the brain were performed in all subjects at screening and then only in subjects with known brain metastasis after randomization on the same schedule as the CAP assessments. Ambiguous brain CT results were confirmed by MRI and brain metastasis had to be treated and stable for at least three months before randomization to meet eligibility requirements. Tumor assessments were continued on the protocol-defined schedule regardless of whether study treatment was given, reduced, held, or discontinued. The same imaging modalities used at screening were to be used for subsequent tumor assessments after randomization. Technetium bone scans (TBS) were also performed in all subjects at screening. After randomization, bone scans were performed only in subjects with known bone metastasis every 16 weeks (± 7 days) throughout the first 12 months on study and then every 24 weeks (± 14 days). Bone scans were used by the Investigator to direct corroborative imaging with CT/MRI if necessary (these CT/MRI findings were used for RECIST 1.1 evaluation). Bone scan findings alone were not used for the determination of progression per RECIST 1.1.

End of radiographic tumor assessments by CT/MRI: For subjects who discontinued study treatment before rPD or within 8 weeks after rPD as determined by the Investigator, final radiographic tumor assessments were performed 8 weeks after rPD (12 weeks for subjects remaining on study treatment for more than one year).

For subjects who continued to receive study drug for more than 8 weeks after Investigator-determined rPD (12 weeks for subjects remaining on study treatment for more than one year), tumor assessments were continued per the protocol-defined schedule until study treatment was permanently discontinued.

Independent Radiology Committee (IRC): For the purpose of determination of the study endpoints of PFS and response rates, a blinded, central review of radiographic images was conducted by an IRC. All radiographic tumor assessments were sent to the central IRC, which also reviewed prior radiation history data for the purpose of selection of target lesions. Details are in the XL184-308 IRC Charter.

For OS, survival status was determined at scheduled visits and every 8 weeks (± 7 days) after the Post-Treatment Follow-up Visit. Subjects were followed until death, consent withdrawn, or Sponsor decision to no longer collect these data.

APPENDIX B: Concordance between IRC and Investigator Assessment of Radiographic Progressive Disease

Table summarizes concordance and discordance between the IRC- and the investigator-determined assessments of PD status (yes vs no). In the PITT population, IRC and investigator agreed on subjects' radiographic PD status 83% of the time for the cabozantinib arm and 77% of the time for the everolimus arm. In the ITT population, they agreed 81% of the time for the cabozantinib arm and 78% of the time for the everolimus arm.

Table summarizes concordance between the IRC and investigator determined assessments of PD dates through the 22 May 2015 data cutoff date for the PITT population. In the PITT population, when both IRC and investigator agreed on PD, they agreed on the dates of PD 42% of the time for the cabozantinib arm and 52% of the time for the everolimus arm. In the ITT population, they agreed on the dates of PD 45% of the time for the cabozantinib arm and 55% of the time for the everolimus arm.

Table 26: Concordance between IRC and Investigator Read in Progressive Disease

		Cabozantinib N = 187 (PITT), N = 330 (ITT)			**Everolimus N = 188 (PITT), N = 328 (ITT)**		
		IRC Read, n (%)					
	Investigator Read, n (%)	Yes	No	Total	Yes	No	Total
PITT Population	Yes	111 (60)	20 (11)	131 (71)	110 (62)	24 (13)	134 (75)
	No	12 (6.5)	42 (23)	54 (29)	17 (9.6)	27 (15)	44 (25)
	Total	123 (66)	62 (34)	185 (100)	127 (71)	51 (29)	178 (100)
ITT Population	Yes	157 (49)	35 (11)	192 (60)	187 (60)	40 (13)	227 (73)
	No	26 (8.1)	103 (32)	129 (40)	29 (9.3)	56 (18)	85 (27)
	Total	183 (57)	138 (43)	321 (100)	216 (69)	96 (31)	312 (100)

Percentages are calculated using the number of subjects with nonmissing status as the denominator.
[Source: CSR Table 26 and Reviewer's Analysis]

Table 27: Concordance between IRC and Investigator Read in Date of Progressive Disease for Tumor Assessments Among Subjects Who Progressed in the PITT Population

	Concordance	**Cabozantinib N = 111**	**Everolimus N = 110**	**Total N = 221**
PITT Population	Yes	47 (42)	57 (52)	104 (47)
	No	64 (58)	53 (48)	117 (53)
		Cabozantinib N = 157	**Everolimus N = 187**	**Total N = 344**
ITT Population	Yes	71 (45)	102 (55)	173 (50)
	No	86 (55)	85 (45)	171 (50)

[Source: CSR Table 27 and Reviewer's Analysis]

<u>Reviewer's Comment:</u> Results for the ITT population were from our analysis. Data was available to produce concordance tables for the ITT population but results were not explicitly presented in the CSR. Results in both populations were consistent.

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

/s/

JOYCE H CHENG
04/14/2016

SHENGHUI TANG
04/14/2016

RAJESHWARI SRIDHARA
04/14/2016

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 #:	NDA208692/S-0000 (IND [redacted] and [redacted]
Drug Name:	Cabometyx™ (cabozantinib), tablets, 20, 40, 60 mg
Indication(s):	The treatment of advanced renal cell carcinoma (RCC) in patients who have received [redacted] prior therapy
Applicant:	Exelixis, Inc. 210 E. Grand Avenue, South San Francisco, CA 94080, USA Laboratory: [redacted]
Date(s):	Received 10/12/2015
Documents Reviewed:	Study XL184-NC-042 (mice) reports and electronic datasets submitted with the electronic submission on 11/13/2015 (via S-004).
Review Priority:	Priority Review
Biometrics Division:	Division of Biometrics VI
Statistical Reviewer:	Feng Zhou, M.S.
Concurring Reviewers:	Karl Lin, Ph.D. Team Leader
Medical Division:	Division of Hematology and Oncology Products
Pharmacology Team:	Eias Zahalka, Ph.D; Todd Palmby, Ph.D
Project Manager:	Rajesh Venugopal, MPH, MBA
Keywords:	Carcinogenicity, Dose response

Table of Contents

1 Summary

This review evaluates statistically the tumorigenicity data of a 26-week intravenous carcinogenicity study of XL184 in Tg.rasH2 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 incidence between each control group and each of the treated groups. There was no difference between the vehicle control and water control based on the survival analysis and tumor analysis. From the statistical point of view, this review concludes that XL184 at high dose decreased survival in male mice. The tumor analysis did not show any statistically significant dose-response relationship in tumor incidence in either sex.

Mouse Study: Mice (25/sex/dose) were dosed by oral gavage with XL184 once every 2 weeks for up to 26 weeks. The XL184 doses were 0, 0, 2, 5, or 15 mkd of XL184 for the vehicle control (VC), water control (WC), low (LD), mid (MD) and high-dose (HD) groups, respectively, in both sexes. Dosing volume was 5 mL/kg in each group. Another group, which included 20 mice for each sex, was dosed by the intraperitoneal injection with a single-dose of 75-mkd N-methyl-N-nitrosourea with a volume of 10 mL/kg on Day 1 to serve as the positive control (PC).

Survival analysis showed that the HD treated male group had a statistically significant increase in mortality when compared to the VC ($p=0.0031$), WC ($p=0.0031$). The trend test showed a statistically significant dose response in mortality among all the treatment groups excluding the PC group in males ($p<0.0001$). The mortality was unaffected in females. The respective survival rates in the VC, WC, LD, MD, HD, PC groups at the termination (Week 27) were 96%, 96%, 100%, 96%, 64%, and 4% in males and 88%, 92%, 80%, 100%, 88%, and 25% in females, respectively. The PC groups in both sexes showed statistically significant increases in mortality when compared with the individual control ($p < 0.001$).

Results of the tumor analysis showed no statistically significant dose response in tumor incidence among the treatment groups excluding the PC group or pairwise difference in tumor incidence in any organs between the individual controls and XL184 treated groups in male and female mice. The PC group showed statistically significant increases in the incidence of a number of tumors in both males and females ($p<0.05$), when compared to the individual controls. Those tumor types included malignant M-Lymphoma in the whole body, squamous B-papilloma in Skin, and squamous B-papilloma and M-papilloma in Stomach.

2 Background

Cabozantinib was granted fast track designation (on 4/8/2015) and breakthrough therapy designation (on 8/21/2015) for the treatment of advanced renal cell carcinoma (RCC) in patients who have received (b) (4) prior therapy indication. The drug product in this NDA is an immediate release (IR) film-coated tablet formulation. This NDA included the nonclinical study (XL184-NC-042), which conducted as Post-Marketing Requirements (PMR, 1970-1) to NDA 203,756 for Cometriq® (cobzantinib, capsules) which was approved on 11/29/2012 for the treatment of progressive, metastatic medullary thyroid cancer (MTC) indication. The final study report was submitted on 7/6/2015 (b) (4), no electronic dataset was submitted. The Information Request (IR) letter was sent out on 11/10/2015. This study evaluated the carcinogenic potential of the test

article, cabozantinib S-malate (hereafter, referred to as XL184), when administered daily by oral gavage to 001178-T (hemizygous) Tg.rasH2 mice for at least 26 weeks.

This reviewer analyzed the SAS data sets of this study received on 11/13/2015 via submission NDA 208,692/S004. The statistical evaluation of survival data and tumor incidence included in the sponsor's report was performed by (b) (4) (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. Eias Zahalka.

3 Mouse Study

Study Report: XL184-NC-042.pdf; SAS data: tumor.xpt

This study assessed the carcinogenic potential of XL184 in male and female Tg.rasH2 mice. The test material was administered at doses of 2, 5, or 15 mkd of XL184 once every 2 weeks by oral gavage for at least 26 weeks. This review refers these dose groups as the low (LD), mid (MD), and high (HD) dose groups, respectively. There are two controls; vehicle control (VC) [EtOH and PEG] and water control (WC) [reverse osmosis (RO)] which will be refers to VC and WC respectively in this review. Dosing volume was 5 mL/kg in each group. There was a positive control (PC) which was dosed with one intraperitoneal dose of N-methyl-N-nitrosourea on Day 1 at a dose level of 75 mg/kg/day and a dose volume of 10 mL/kg. There were 25 mice for each sex and dose group except positive control group which included 20 mice for each sex.

Parameters evaluated included survivability, clinical observations, body weights, food consumption, and clinical and anatomic pathology. Carcinogenicity assessment was based on mice palpation to monitor for masses, survival data, and anatomic pathology.

For all the sponsor's analyses, the data from the positive control group were excluded and actual dose levels were used in the statistical analysis.

3.1 Sponsor's Analyses

3.1.1 Survival Analysis

The sponsor performed tests to compare survival with a two sided risk for increasing and decreasing mortality with dose; and tests for dose response (the water control was excluded from this test) and for each treated group against vehicle control using Kaplan-Meier product-limit estimation curves, along with log-rank and Wilcoxon tests. These were performed using the LIFETEST procedure in SAS. The time to death or sacrifice (in weeks) was the dependent variable. Treatment group was included as the strata. Animals with a death or sacrifice status recorded as a planned sacrifice (interim or terminal) or an accidental death were censored in the analysis.

Sponsor's findings: The sponsor reported that the male high dose group (15 mg/kg/day) had higher mortality than the vehicle control group (9/25 versus 1/25), with p=0.0046 and p=0.0043 for the Log-Rank and Wilcoxon tests respectively. The trend was also significant ($p < 0.0001$ for both the Log-

Rank and Wilcoxon tests). No other significant findings were noted in males and in females.

The Sponsor's Report of Total Unscheduled Deaths (n=25/sex/dose group)

Males

		Unadjusted Survival Incidence Rate				
Group	Trend	1	2	3	4	5
Dose Level(mg/kg/day)	(1,3,4,5)	0	0	2	5	15
Group Size		25	25	25	25	25
TK		24	24	25	24	16
Deaths		1	1	0	1	9
Log-Rank P-value (v2)	<.0001**	N/A	0.9885	0.3173	0.9885	0.0046**
Wilcoxon P-value (v2)	<.0001**	N/A	0.9772	0.3173	0.9772	0.0043**

Females

		Unadjusted Survival Incidence Rate				
Group	Trend	1	2	3	4	5
Dose Level(mg/kg/day)	(1,3,4,5)	0	0	2	5	15
Group Size		25	25	25	25	25
TK		22	23	20	25	22
Deaths		3	2	5	0	3
Log-Rank P-value (v2)	0.1536	N/A	0.6544	0.4268	0.0770	0.9792
Wilcoxon P-value (v2)	0.1466	N/A	0.6683	0.4118	0.0771	0.9589

* = Significant at 5% level; ** = Significant at 1% level.
Dose groups 1=VC, 2=WC, 3=LD, 4=MD, and 5=HD.
[Source: page 1242 and page 1243 of study report of XL184-NC-042.pdf]

3.1.2 Tumor Data Analysis

The sponsor stated that only tumors from tissues that were listed in the protocol to be examined for all animals were analyzed. For each given tumor type, the statistical analysis was performed if the incidence in at least one group (dosed and/or water control) was increased by at least two occurrences over the vehicle control group. The sponsor analyzed the tumor incidence data with a one sided risk for increasing incidence with dose. The water and vehicle controls were compared with two sided risk. Tests were performed for dose response (the water control was excluded from this test) and for each treated group against vehicle control.

Occult or non-palpable tumors were analyzed by the IARC asymptotic fixed interval based prevalence test (Peto et al., 1980)[2]. The cut off points for the interval based test were Weeks 1-13, 14-26, and terminal sacrifice. Fatal and non-fatal tumors were analyzed together, with separate strata for each. There were no tumors of uncertain context. The test was implemented using PROC MULTTEST in the SAS system. In the case of sparse tables (<10 total in the groups being analyzed for the trend or pairwise test), the exact form of the test was used. Otherwise, the asymptotic version of the test was used. Animals were assigned to the terminal sacrifice strata based on the death or sacrifice status recorded in the data and were not be assigned based on their week of necropsy.

Unadjusted P-values were reported for tumors. The indication of a possible treatment effect was assessed on the basis of rare or common tumor type, in line with the current FDA guidelines (FDA Draft Guidance for Industry, 2001)[1]. The incidence rate for defining whether a tumor type is rare or common is based on site specific background historical data. The study pathologist determined whether a tumor type was rare or common.

Sponsor's findings: No statistically significant findings in males or females were observed.

File Name: outputfile1988218205.pdf

Results of Statistical Analyses of Neoplastic Lesions – Males

			Unadjusted Lifetime Incidence Rate				
	Group	Trend	1	2	3	4	5
Tissue and Lesion	Dose Level(mg/kg/day)	(1,3,4,5)	0	0	2	5	15
Body, Whole/Cavity							
M-Hemangiosarcoma							
	No. Examined		25	25	25	25	25
	Fatal Tumor		0	1	0	1	0
	Incidental Tumor		2	1	3	0	4
	Total Tumors		2	2	3	1	4
	Peto P-value (v2)	0.0812	N/A	N/A	N/A	N/A	0.1603

[Source: page 1244 of study report of XL184-NC-042.pdf]

Results of Statistical Analyses of Neoplastic Lesions – Females

			Unadjusted Lifetime Incidence Rate				
	Group	Trend	1	2	3	4	5
Tissue and Lesion	Dose Level(mg/kg/day)	(1,3,4,5)	0	0	2	5	15
Body, Whole/Cavity							
M-Hemangiosarcoma							
	No. Examined		25	25	25	25	25
	Fatal Tumor		1	1	1	0	0
	Incidental Tumor		1	3	2	2	2
	Total Tumors		2	4	3	2	2
	Peto P-value (v2)	N/A	N/A	0.6792	N/A	N/A	N/A
Lung							
M-Carcinoma, bronchiolo-alveolar							
	No. Examined		25	25	25	25	25
	Fatal Tumor		0	0	0	0	1
	Incidental Tumor		0	0	0	0	1
	Total Tumors		0	0	0	0	2
	Peto P-value (v2)	0.0638	N/A	N/A	N/A	N/A	0.2553

[Source: page 1245 of study report of XL184-NC-042.pdf]

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 submission NDA 208,692 on 11/13/2015 (via S-0004).

3.2.1 Survival Analysis

The Kaplan-Meier curves for survival rates of all treatment groups are given in Figures 1A and 1B in the appendix for male and female mice, respectively. The intercurrent mortality data of all treatment groups are given in Tables 1A and 1B 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 2A and 2B in the appendix for male and female mice, respectively.

File Name: outputfile1988218205.pdf

Reviewer's findings: Survival analysis results showed that the HD treated male group had a statistically significant increase in mortality when compared to the VC (p=0.0031), WC (p=0.0031). The trend test showed a statistically significant dose response in mortality among all the treatment groups excluding the PC group in males (p<0.0001). The mortality was unaffected in females. The respective survival rates in the VC, WC, LD, MD, HD, PC groups at the termination (Week 27) were 96%, 96%, 100%, 96%, 64%, and 4% in males and 88%, 92%, 80%, 100%, 88%, and 25% in females respectively . The PC groups in both sexes showed statistically significant increases in mortality when compared with the individual controls (p < 0.001).

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 in a treatment group 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 in the treatment group that dies at week w_h without developing the tumor before the end of the study gets a score of $s_h = \left(\frac{w_h}{w_{\max}} \right)^k$ <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 the given tumor type being tested, 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. The tumor rates and the p-values of the tested tumor types are listed in Tables 3A and 3B in the appendix for male and female mice, respectively. The tumor rates and the p-values of the comparisons between the vehicle control and positive control are listed in Tables 4A and 4B in the appendix for male and female mice, respectively.

Multiple testing adjustment: For the adjustment of multiple testing of dose response relationship, the FDA guidance for the 26 weeks transgenic mouse study design and data analysis suggests the use of test levels α=0.05 for both the trend tests and the pairwise comparisons regardless a tumor type is common or rate.

Reviewer's findings: Based on the criteria of adjustment for multiple testing discussed above, results of the tumor analysis showed no statistically significant dose response in tumor incidence among the treatment groups excluding the PC group or pairwise difference in tumor incidence in any organs between the individual controls and XL184 treated groups in male and female mice. The PC group showed statistically significant increases in the incidence of a number of tumors in both males and females (p<0.05), when compared to the individual controls. Those tumor types included malignant M-Lymphoma in the whole body, squamous B-papilloma in Skin, and squamous B-papilloma and M-papilloma in Stomach.

Tumor Types with P-Values ≤ 0.05 for Pairwise Comparisons of Vehicle Control in Female Mice

Organ Name	Tumor Name	0 mkd	2 mkd	5 mkd	15 mkd	P-Value

File Name: outputfile1988218205.pdf

		VC N=25	LD N=25	MD N=25	HD N=25	Dose Response	VC vs. LD	VC vs. MD	VC vs. HD
Lung	M-Carcinoma, bronchi	0	0	0	2	0.0645	.	.	0 2553

Tumor Types with P-Values ⩽ 0.05 for Pairwise Comparisons of Water Control in Female Mice

Organ Name	Tumor Name	0 mkd WC N=25	2 mkd LD N=25	5 mkd MD N=25	15 mkd HD N=25	P-Value			
						Dose Response	VC vs. LD	VC vs. MD	VC vs. HD
Lung	M-Carcinoma, bronchi	0	0	0	2	0.0645	.	.	0 2553

Tumor Types with P-Values ≤ 0.05 for Pairwise Comparisons between VC, WC, and PC in Male Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	0 mg/kg/day W(N=25)	75 mg/kg/day PC (N=20)	P-Value VC vs. WC	P-Value VC vs. PC	P-Value WC vs. PC
Body, Whole/Cav	M-Lymphoma, malignan	0	0	7	.	<0.001*	<0.001*
Skin/Subcutis	B-Papilloma, squamou	0	0	16	.	<0.001*	<0.001*
Stomach, Nongla	B-Papilloma, squamou	0	0	10	.	<0.001*	<0.001*
	M-Carcinoma, squamou	1	0	6	0.5000	0.0035*	<0.001*

Tumor Types with P-Values ≤ 0.05 for Pairwise Comparisons between VC, WC, and PC in Female Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	0 mg/kg/day W(N=25)	75 mg/kg/day PC (N=20)	P-Value VC vs. WC	P-Value VC vs. PC	P-Value WC vs. PC
Body, Whole/Cav	M-Lymphoma, malignan	0	0	8	.	<0 001*	<0.001*
Skin/Subcutis	B-Papilloma, squamou	0	0	8	.	<0 001*	<0.001*
Stomach, Nongla	B-Papilloma, squamou	0	0	12	.	<0 001*	<0.001*

4 Conclusion

This review evaluates statistically the tumorigenicity data of a 26-week intravenous carcinogenicity study of XL184 in Tg.rasH2 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 incidence between each control group and each of the treated groups. There was no difference between the vehicle control and water control based on the survival analysis and tumor analysis. From the statistical point of view, this review concludes that XL184 at high dose decreased survival in male mice. The tumor analysis did not show any statistically significant dose-response relationship in tumor incidence in either sex.

The PC group showed statistically significant increases in the incidence of a number of tumors in both males and females ($p<0.05$), when compared to the individual controls. Those tumor types included malignant M-Lymphoma in the whole body, squamous B-papilloma in Skin, and squamous B-papilloma and M-papilloma in Stomach.

Feng Zhou, M.S.
Mathematical Statistician

File Name: outputfile1988218205.pdf

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

cc:
Dr. Eias Zahalka
Dr. Todd Palmby
Dr. Tsong
Ms. Zhou
Dr. Lin
Ms. Patrician

5 Appendix

Table 1A: Intercurrent Mortality – Male Mice

	VC 0 mkd (n=25)		WC 0 mkd (n=25)		LD 2 mkd (n=25)		MD 5 mkd (n=25)		HD 15 mkd (n=25)		PC 75 mkd (n=20)	
Week	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %
0 – 26	1	4.00	1	4 00	0	0	1	4 00	9	36.00	16	80.00
Ter. Sac.	24	96.00	24	96.00	25	100.00	24	96.00	16	64.00	4	20.00

[#] Cum. %: Cumulative percentage except for Ter. Sac.

Table 1B: Intercurrent Mortality - Female Mice

	VC 0 mkd (n=25)		WC 0 mkd (n=25)		LD 2 mkd (n=25)		MD 5 mkd (n=25)		HD 15 mkd (n=25)		PC 75 mkd (n=20)	
Week	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %	No. Death	Cum. %
0 – 26	3	12.00	2	8 00	5	20.00	0	0	3	12.00	15	75.00
Ter. Sac.	22	88.00	23	92.00	20	80.00	25	100.00	22	88.00	5	25.00

Table 2A: Intercurrent Mortality Comparison – Male Mice

Test	Statistic	P-Value Compared to VC	P-Value Compared to WC	P-Value Compared to PC
Dose-Response	Likelihood Ratio	<0.0001	<0 0001	<0 0001
Homogeneity	Log-Rank	<0.0001	<0 0001	<0 0001
High Dose-	Likelihood Ratio	0.0031	<0 0001	<0 0001
Homogeneity	Log-Rank	0.0046	<0 0001	<0 0001

Table 2B: Intercurrent Mortality Comparison – Female Mice

Test	Statistic	P-Value Compared to VC	P-Value Compared to WC	P-Value Compared to PC
Dose-Response	Likelihood Ratio	0.7282	0.9617	<0 0001
Homogeneity	Log-Rank	0.1536	0.1176	<0.0001

Table 3A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Vehicle Control – Male Mice

Organ Name	Tumor Name	0 mkd VC N=25	2 mkd LD N=25	5 mkd MD N=25	15 mkd HD N=25	P-Value			
						Dose Response	VC vs. LD	VC vs. MD	VC vs. HD
Body, Whole/Cav	M-Hemangiosarcoma	2	3	1	4	0.1161	0.5000	0.5000	0.2308
	M-Lymphoma, malignan	0	0	0	1	0.2128	.	.	0.4444
Lung	B-Adenoma, bronchiol	1	2	1	0	0.7883	0.5000	0.7449	0.4444
	M-Carcinoma, bronchi	2	1	0	0	0.9269	0.5000	0.7449	0.6970
Muscle, Skeleta	M-Sarcoma	0	1	0	0	0.4681	0.5000	.	.
Stomach, Nongla	M-Carcinoma, squamou	1	0	0	0	0.7340	0.5000	0.4898	0.4444

Table 3B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Water Control – Male Mice

Organ Name	Tumor Name	0 mkd WC N=25	2 mkd LD N=25	5 mkd MD N=25	15 mkd HD N=25	P-Value			
						Dose Response	WC vs. LD	WC vs. MD	WC vs. HD
Body, Whole/Cav	M-Hemangiosarcoma	2	3	1	4	0.1161	0 5000	0.5000	0.2308
	M-Lymphoma,	0	0	0	1	0.2128	.	.	0.4444

File Name: outputfile1988218205.pdf

Organ Name	Tumor Name	0 mkd WC N=25	2 mkd LD N=25	5 mkd MD N=25	15 mkd HD N=25	P-Value			
						Dose Response	WC vs. LD	WC vs. MD	WC vs. HD
	malignan								
Liver	B-Adenoma, hepatocel	1	0	0	0	0.7340	0 5000	0.4898	0.4444
Lung	B-Adenoma, bronchiol	2	2	1	0	0.8923	0 6954	0.4844	0.6970
	M-Carcinoma, bronchi	0	1	0	0	0.4681	0 5000	.	.
Muscle, Skeleta	M-Sarcoma	0	1	0	0	0.4681	0 5000	.	.

Table 4A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Vehicle Control – Female Mice

Organ Name	Tumor Name	0 mkd VC N=25	2 mkd LD N=25	5 mkd MD N=25	15 mkd HD N=25	P-Value			
						Dose Response	VC vs. LD	VC vs. MD	VC vs. HD
Body, Whole/Cav	M-Hemangiosarcoma	2	3	2	2	0.5587	0.4782	0.3369	0.6961
	M-Lymphoma, malignan	0	1	0	0	0.5161	0.4889	.	.
Lung	B-Adenoma, bronchiol	2	0	0	0	0.9396	0.7326	0.7757	0.7556
	M-Carcinoma, bronchi	0	0	0	2	0.0645	.	.	0.2553
Tongue	M-Carcinoma, squamou	0	1	0	0	0.5161	0.4889	.	.

Table 4B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons with Water Control – Female Mice

Organ Name	Tumor Name	0 mkd WC N=25	2 mkd LD N=25	5 mkd MD N=25	15 mkd HD N=25	P-Value			
						Dose Response	WC vs. LD	WC vs. MD	WC vs. HD
Body, Whole/Cav	M-Hemangiosarcoma	4	3	2	2	0.7656	0.4490	0.6864	0.6460
	M-Lymphoma, malignan	0	1	0	0	0.5161	0.4889	.	.
Harderian Gland	B-Adenoma	1	0	0	0	0.7500	0.4773	0.5208	0.5000
Kidney	B-Adenoma, tubule ce	1	0	0	0	0.7500	0.4773	0.5208	0.5000
	M-Carcinoma, tubule	1	0	0	0	0.7500	0.4773	0.5208	0.5000
Lung	B-Adenoma, bronchiol	1	0	0	0	0.7500	0.4773	0.5208	0.5000
	M-Carcinoma, bronchi	0	0	0	2	0.0645	.	.	0.2553
Tongue	M-Carcinoma, squamou	0	1	0	0	0.5161	0.4889	.	.

Table 5A: Tumor Rates and P-Values for Comparisons between WC, VC, and PC – Male Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	0 mg/kg/day WC (N=25)	75 mg/kg/day PC (N=20)	P-Value VC vs. WC	P-Value VC vs. PC	P-Value WC vs. PC
Body, Whole/Cav	M-Hemangiosarcoma	2	2	2	0.6954	0 3191	0.3191
	M-Lymphoma, malignan	0	0	7	.	<0.001*	<0.001*
Liver	B-Adenoma, hepatocel	0	1	0	0.5000	.	0.2647
Lung	B-Adenoma, bronchiol	1	2	0	0.5000	0 2647	0.4652
	M-Carcinoma, bronchi	2	0	0	0.7551	0.4652	.
Skin/Subcutis	B-Papilloma, squamou	0	0	16	.	<0.001*	<0.001*
	M-Carcinoma, squamou	0	0	1	.	0 2857	0.2857
Stomach, Nongla	B-Papilloma, squamou	0	0	10	.	<0.001*	<0.001*
	M-Carcinoma, squamou	1	0	6	0.5000	0.0035*	<0.001*

*Indicted the significant at 0.05 alpha levels.

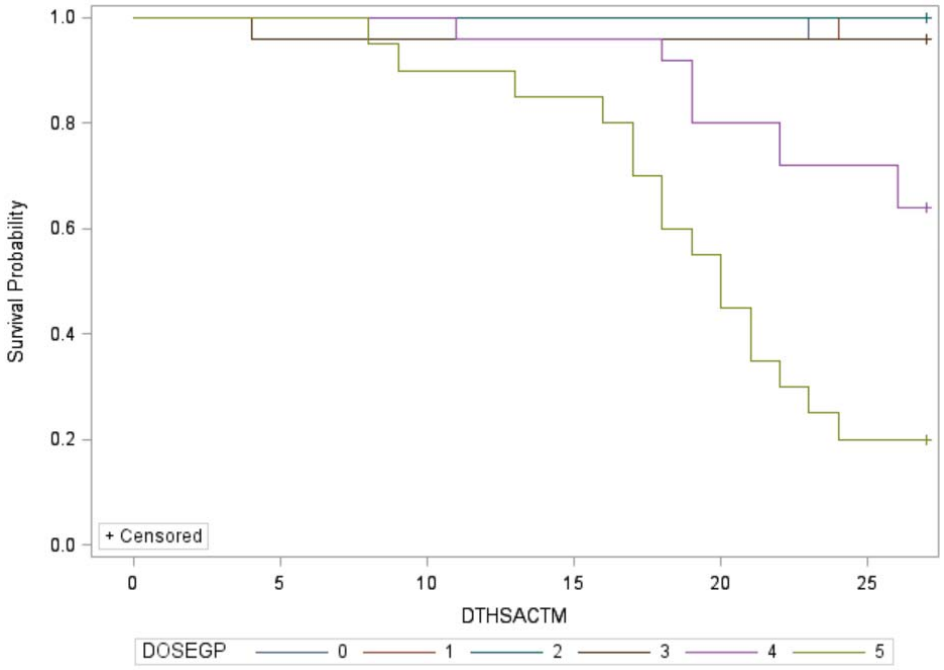
File Name: outputfile1988218205.pdf

Table 5B: Tumor Rates and P-Values for Comparisons between WC, VC, and PC – Female Mice

Organ Name	Tumor Name	0 mg/kg/day VC (N=25)	0 mg/kg/day WC(N=25)	75 mg/kg/day PC (N=20)	P-Value VC vs. WC	P-Value VC vs. PC	P-Value WC vs. PC
Body, Whole/Cav	M-Hemangiosarcoma	2	4	0	0.3540	0 5490	0.7971
	M-Lymphoma, malignan	0	0	8	.	<0.001*	<0 001*
Harderian Gland	B-Adenoma	0	1	0	0.5000	.	0.3235
Kidney	B-Adenoma, tubule ce	0	1	0	0.5000	.	0.3235
	M-Carcinoma, tubule	0	1	0	0.5000	.	0.3235
Lung	B-Adenoma, bronchiol	2	1	0	0.5000	0 5490	0.3235
	M-Carcinoma, bronchi	0	0	1	.	0 3429	0.3429
Mammary Gland,	M-Carcinoma	0	0	1	.	0 3429	0.3429
Ovary	M-Malignant granulos	0	0	1	.	0 3235	0.3235
Skin/Subcutis	B-Papilloma, squamou	0	0	8	.	<0.001*	<0 001*
	M-Carcinoma, squamou	0	0	2	.	0 0980	0.0980
Stomach, Nongla	B-Papilloma, squamou	0	0	12	.	<0.001*	<0 001*
	M-Carcinoma, squamou	0	0	1	.	0 3235	0.3235

*Indicted the significant at 0.05 alpha levels.

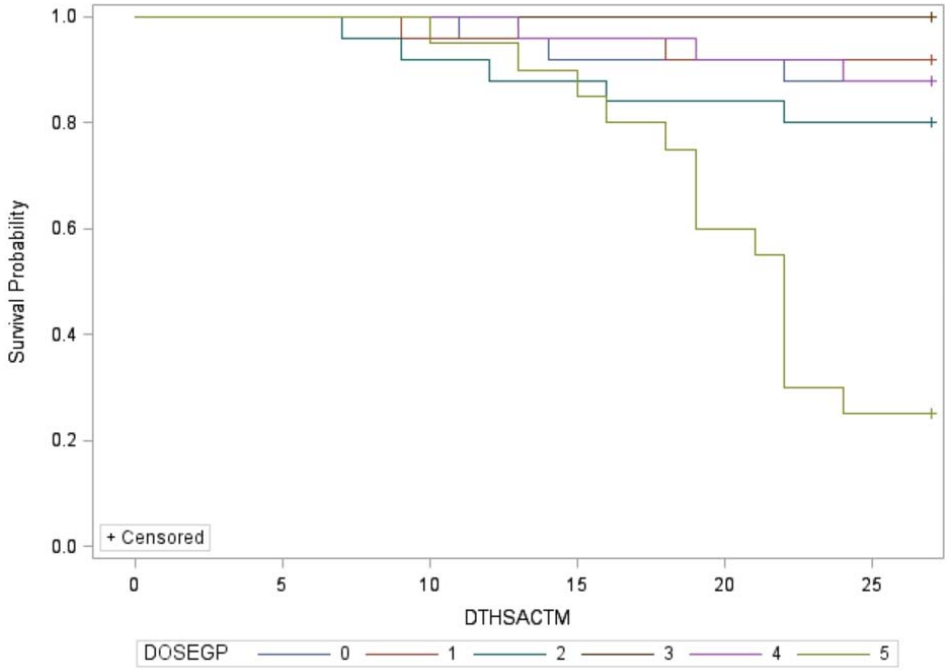
Figure 1A: Kaplan-Meier Survival Functions for Male Mice



Note: dose group should be 0=VC, 1=WC, 2=2, 3=5, 4=15-mg/kg/day of XL184, or 5=75-mg/kg/day of N-methyl-N-nitrosourea (PC)

File Name: outputfile1988218205.pdf

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



Note: dose group should be 0=VC, 1=WC, 2=2, 3=5, 4=15-mg/kg/day of XL184, or 5=75-mg/kg/day of N-methyl-N-nitrosourea (PC)

6 References

1. 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.
2. 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", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva,* 311-426, 1980.

--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
03/07/2016

KARL K LIN
03/08/2016
Concur with review

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 208692 **Applicant:** Exelixis **Stamp Date:** 12/22/15

Drug Name: Cabozantinib **NDA/BLA Type:** NDA

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

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

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? _yes____

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

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

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

File name: 5_Statistics Filing Checklist for NDA_208692

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

/s/

JOYCE H CHENG
01/20/2016

SHENGHUI TANG
01/21/2016