CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

203756Orig1s000

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 #:	203756
Supplement #:	SN0002
Drug Name:	Cabozantinib (S)-malate (Cometriq)
Indication(s):	Thyroid cancer (MTC)
Applicant:	Exelixis
Date(s):	May 21, 2012
Review Priority:	Priority
Biometrics Division:	V
Statistical Reviewer:	Yuan-Li Shen
Concurring Reviewers:	Kun He
	Rajeshwari Sridhara
Medical Division:	Oncology Products 2
Clinical Team:	Ruthann Giusti
	Suzanne Demko
	Patricia Keegan
Project Manager:	Gina Davis

Keywords:
Link to keywords: a log-rank test, hazard ratio, Cox model

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

The applicant submitted Study XL184-301 to seek an indication for the treatment of patients with progressive, unresectable locally advanced or metastatic medullary thyroid cancer (MTC).

Based on study XL184-301, the results demonstrated statistically significant improvement based on the progression free survival in favor of the Cabozantinib treatment (the median PFS time was 11.2 months [95% CI=8.4, 13.7] in the Cabozantinib arm and 4 months [95% CI=3.0, 5.4] in the placebo arm; the hazard ratio estimate was 0.28 [95% CI=0.19, 0.40]). The favorable results from the Cabozantinib arm were robust based on various sensitivity analyses and consistent across different subgroups (including RET mutation status). The result based on the objective response rate also demonstrated statistically significant improvement in favor of the Cabozantinib arm (ORR=27% vs. 0% for Cabozantinib arm vs. placebo, respectively). However, based on 44% information level, the result did not demonstrate treatment benefit for the overall survival (HR=0.997, 95%=0.64, 1.54). Based on the 120-day updated analysis of OS (i.e. 75% information level), the result still did not show beneficial effect for the Cabozantinib treated arm (HR=0.825, 95% CI=0.598, 1.140) with median OS time 26.02 [95% CI=22.90, 30.72] and 20.34 [95% CI=16.39, 26.68] for the Cabozantinib arm and placebo arm, respectively).

In conclusion, this statistical reviewer confirms the applicant's results submitted. Whether the results demonstrate an overall favorable benefit to risk ratio in supporting an indication of the Cabozantinib treatment in patients with progressive, unresectable locally advanced or metastatic medullary thyroid cancer will defer to the clinical review team.

2 INTRODUCTION

2.1 Overview

The applicant proposed the following indication statement in the patient package insert:

COMETRIQ is indicated for the treatment of patients with progressive, unresectable locally advanced or metastatic medullary thyroid cancer (MTC).

One phase III study, XL184-301 was used to support this application. Study XL184-301 was an international, randomized, double-blinded, phase 3 efficacy study in subjects with unresectable, locally advanced, or metastatic medullary thyroid cancer. This study received a Special Protocol Assessment agreement (SPA) on June, 2008. The drug received orphan drug designation on November, 2010. The primary objective of this study was to evaluate progression-free survival (PFS) with Cabozantinib (XL184) treatment as compared with placebo. The secondary objectives of the study were to evaluate the overall survival (OS) and objective response rate (ORR). The first patient was enrolled on 9/10/2008 and the study was on-going at the time of submission. There were 98 principal investigators who enrolled subjects at 90 unique sites in 23 countries in Europe, North America, Middle East, South America, and Asia.

Some key information for the supporting study is summarized in the following table:

Table 1 Summaries of the Key Information for the Supporting Phase III tries

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
XL184-301	Randomized, double-blind, phase III study	Treatment Period: Each cycle of the treatment Period includes 4 weeks of daily administration of XL184 or placebo.	Follow-up quarterly (12 weeks ± 15 days) after the 30-days post dosing assessment. The median follow-up time : XL184 : 8.4 months Placebo : 7.8 months	XL184 : 219 Placebo : 111	Patients with unresectable, locally advanced, or metastatic medullary thyroid cancer.

2.2 Data Sources

The application's data (including raw and analysis datasets) from the original submission for study XL184-301 is located in the following link: \\Cdsesub5\evsprod\NDA203756\0002\m5\datasets\xl184-301.

The clinical study reports and the statistical analysis plan for this study are located in the following link:

\\Cdsesub5\evsprod\NDA203756\0002\m5\53-clin-stud-rep\535-rep-effic-safety-stud\medullary-thyroid-cancer\5351-stud-rep-contr\xl184-301.

On 8/9/2012, the applicant responded to the agency's request with regard to the data for the biomarker (Calcitonin and CEA) and provided SAS exported datasets for further evaluation. This biomarker data is located in the following link:

\\Cdsesub1\evsprod\NDA203756\0012\m5\datasets\xl184-301-cbm-001\analysis\legacy\datasets

The SAS programs that were used to derive the analysis datasets and perform the analysis were also included in the link shown above.

3 STATISTICAL EVALUATION

The original protocol for study XL184-301 was finalized on 4/21/2008 and subsequently undergone two amendments. The items that were revised and may affect the efficacy evaluation are listed below:

Amendment 1 (6/11 2008)

- Subject stratification categories were modified to the following categories:
  - Age : ≤ 65 years vs >65 years
  - Known prior receipt of a TKI (Yes versus No)
- The enrollment of subjects with only nonmeasurable disease was limited to 31 (10% of the total number of subjects) to minimize heterogeneity in the assessment of tumor response.

Amendment 2 (9/24/2010)

- For study eligibility, PD was determined by the investigator at screening instead of requiring confirmation of PD by a blinded radiologist at the IRC.

The final draft statistical analysis plan (SAP) for Special Protocol Assessment (SPA) was dated 4/21/2008. There were a few SAP revisions that might affect the efficacy analyses as shown below:

7/22/2008: Revised stratification factors to match Protocol Amendment 1.0;

2/7/2011 : Clarified the method used to address missing or inadequate tumor assessments for PFS and ORR per FDA's feedback; Clarified details of planned supportive analyses and added additional sensitivity analysis of PFS based on investigator assessment of radiographic progression.

8/29/2011: An imputation rule for partial start dates of subsequent anticancer therapy was established; clarified censoring dates for patients who took anti-cancer therapy.

Reviewer's Comments:
Since the revisions of the stratification factors (7/22/2008) were before the first patient was enrolled (9/10/2008), the revision would not affect the efficacy analyses.

3.1 Data and Analysis Quality

The applicant submitted CRF and analysis datasets, the defined files for the variables and the corresponding SAS programs for documentation of the analysis results. The documentation for the derived variables appears to be easy to follow. The guide for reviewer documents also provides a good overview of the data (e.g. naming convention for the variables and relationship between the CRF and analysis datasets, etc). The reviewer was able to duplicate the analysis results based on the raw dataset as well as from the analysis datasets.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

Study XL1840301 was an international, randomized (2: 1), double-blind, phase III study of XL184 versus placebo in patients with unresectable, locally advanced, or metastatic medullary thyroid cancer. Patients who were 18 years old or older, had an ECOG performance status ≤2, had documented progression of disease (PD) within 14 months of screening, had recovered to National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v3.0 Grade ≤ 1 from clinically significant AEs due to medications administered prior to randomization and had adequate organ and marrow functions as described in the entry criteria, etc, were eligible to be enrolled. Patients who received prior systemic anti-tumor therapy within 4 weeks of randomization, received radiation to ≥25% of bone marrow, received investigational agents within 4 weeks of randomization, had brain metastases or spinal cord compression, had a history of clinically significant hematemesis or a recent history of hemoptysis of >2.5 mL or red blood or other signs indicative of pulmonary hemorrhage or evidence of endobronchial lesion(s), had urine protein /creatinine ratio of ≥ 1, had serious intercurrent illness, etc, were excluded from the study. For purpose of subject management (including the decision to discontinue study treatment), the protocol was revised in Amendment 2 to revise the requirement of the documented radiologic PD based on investigators' assessment. Prior to Amendment 2, the confirmed PD was determined by the Independent Review Committee (IRC). Since Amendment 2 was implemented after 295/330 (89%) patients enrolled, the majority of the patients had the documented PD determined by the IRC.

Randomization was based on a permuted block design. The stratification factor includes Age (≤ 65 years vs. > 65 years) and prior tyrosine kinase inhibitor status (Known prior receipt of a tyrosine kinase inhibitor=Yes vs. No). Randomization for treatment assignment was performed using an interactive voice (or web) response system (IVRS).

Patients were randomized to receive Cabozantinib 175 mg or placebo in 2:1 ratio.

The study includes the following periods:

- Pre-Treatment Period: screening assessments must be performed within 28 days of randomization.
- Treatment Period: Each cycle of the treatment Period includes 4 weeks of daily administration of XL184 or placebo.

  Note: Treatment continued until a patient had either investigator assessed progression per modified Response Evaluation Criteria in Solid Tumors (mRECIST) [defined in the Independent Review Committee (IRC)], or unacceptable toxicity.

- Post-Treatment Period: patients had an end-of-treatment assessment at 30 days after the last dose of study treatment. The investigator obtained follow-up information quarterly (± 15 days) thereafter.

Tumor assessments were performed approximately every 12 weeks (± 5 days) from randomization until disease progression, as determined by the investigator per mRECIST.

The study remained blinded until the primary PFS analysis was complete. Crossing-over of patients receiving placebo to receive XL184 after progression was not permitted in this study.

The primary objective of this study is to evaluate progression-free survival (PFS). The secondary objectives of this study include:

- To evaluate overall survival (OS);
- To evaluate the objective response rate (ORR) and duration of response per mRECIST;
- To evaluate changes in serum levels of calcitonin and carcinoembryonic antigen (CEA) as prognostic biomarkers;

- To assess the potential relationship between RET germline and/or tumor DNA sequence alteration and the efficacy of XL184;
- To assess the pharmacodynamic effects of XL184;
- To evaluate the safety and tolerability of XL184 treatment;
- To assess the PK of XL184 in a subset of subjects.

The exploratory objective of the study is to evaluate subject self-assessment parameters and symptom burden as per the MD Anderson Symptom Inventory (MDASI) Thyroid Module (THY).

Sample Size Calculation

Assuming exponential distribution of PFS time and proportional hazards, 138 PFS events were required for this study to have a 90% power to detect a hazard ratio (HR) of 0.571 based on a log-rank test and a 2-sided 5% significance level. This effect corresponds to a 43% reduction in the PFS risk, or a difference of 8 months versus 14 months in median PFS time.

Assuming one interim analysis of OS at the 31% information level at the time of the primary analysis of PFS and a subsequent primary analysis, 217 deaths were required to have an 80% power to detect a HR of 0.667 using a log-rank test and a 2-sided 4% significance level. This effect corresponds to a 33.3% reduction in the risk of death, or a treatment difference of 22 versus 33 months in median survival time.

A total of 315 eligible subjects (210 in XL184 and 105 in placebo) were planned to be randomized and followed to observe the required number of events within the planned study duration (33 months accrual; approximately 66 months total to observe the required deaths for OS).

Interim Analysis

One interim analysis was planned at 31% information level (approximately 67 deaths) when the primary efficacy analysis on PFS was performed. If the result of the primary endpoint (PFS) was significant, the two key secondary endpoints would be tested in parallel. The ORR and OS would be tested at 2-sided 0.01 and 0.04 levels, respectively. An interim analysis of OS was expected to be performed at the 0.00006 level per a Lan-DeMets O'Brien-Fleming alpha spending function based on an expected 31% information level. The actual alpha level would be based on the actual information fraction at the time of the analysis. If the result of the interim

analysis for OS was not significant, the primary analysis of OS would be performed when the required number of deaths (217) had been observed. The primary analysis was expected to be conducted at the 0.04 significance level per the alpha spending function.

Primary Efficacy Endpoint

PFS was defined as the time from randomization to the earlier of the following events:

- Documented disease progression (IRC determined per mRECIST)
- Death due to any cause.

The censoring rules for the primary analysis of PFS were listed below:

- Patients who died >26 weeks after their last tumor assessment were right censored to the date of their last tumor assessment;
- Patients who did not die within 26 weeks of randomization and did not have any post-baseline tumor assessments were right censored on the date of randomization;
- Patients who did not have tumor assessment for >26 weeks were right censored at their last tumor assessment before the missing tumor assessments;
- Patients who received subsequent anti-cancer therapy before experiencing an event were censored at the date of the most recent adequate tumor assessment prior to the date of starting anti-cancer therapy;
- Patients who did not experience an event at the time of data cutoff were right censored on the date of their last tumor assessment.

Secondary Efficacy Endpoint

Objective response rate (ORR) and overall survival (OS) were the two key secondary endpoints.

Overall survival was defined as time from randomization to death due to any cause. OS would be censored at the last date known to be alive if a patient was alive at the cut-off time or was lost to follow-up.

ORR was defined as the proportion of patients with measurable disease at baseline who had the best overall response (BOR) of CR or PR at the time of data cutoff which was confirmed by a subsequent visit ≥28 days later. If multiple assessments were performed and an overall response of CR or PR was observed, the latest assessment date within the set would be chosen as the response date.

Supportive Secondary Efficacy Endpoints:

The supportive secondary Efficacy Endpoints are summarized below:

- Disease Stabilization Rate: defined as the proportion of patients for whom the BOR is confirmed CR/PR or SD ≥24 weeks;
- Duration of Response : defined as the time from first documentation of objective response that is subsequently confirmed at a visit that is ≥28 days later to disease progression or death due to any cause;
- Duration of Stable Disease: defined as the number of days among patients with a BOR of SD between the date of randomization and the date of disease progression or death due to any cause.

Efficacy Exploratory Analysis:

The impact of a set of baseline and demographic characteristics on PFS, OS and ORR would be evaluated. These factors include:

- Age (≤ 45 years, age 45-65, age > 65 years);
- Gender (male, female);
- Race (White, Non-white);
- ECOG status (0, ≥1);
- Prior tyrosine kinase inhibitor status (Yes, No, Unknown);
- Geographic region (U.S., Europe, Rest of World);
- RET mutation status (Positive, Negative, Unknown);
- Best response to prior therapy (CR/PR, SD, PD, Unknown);
- Prior anti-cancer regimens (no prior regimens, 1 prior regimen, ≥2 prior regimens);
- Prior radiotherapy (Yes, No);
- Disease stage at current diagnosis (locally-advanced, metastatic);
- Bone metastasis at baseline per IRC (bone metastasis only, bone metastasis and soft tissue metastasis, no bone metastasis).

3.2.2 Patient Disposition, Demographic and Baseline Characteristics

The first patient was enrolled on 9/10/2008 and there are 45% and 14% of the patients still under study treatment for XL184 and placebo arm, respectively, at the time of database cut-off date (6/15/2011).

A total of 330 patients (219 in Cabozantinib and 111 in placebo arm) were randomized to receive treatments. There were a total of 90 unique sites from 23 countries in Europe, North America, Middle East, South America and Asia. The enrollment disposition by regions is shown in the following table:

Table 2 Enrollment Disposition by Regions

	XL184 N=219	Placebo N=111
Europe	124 (57%)	60 (54%)
North America	69 (32%)	33 (30%)
Rest of World	26 (12%)	18 (16%)

There were 55% and 87% of the patients discontinued study treatment in XL184 and placebo arm, respectively. The primary reasons for discontinuation from study treatment were due to disease progression determined per investigator (27% and 60% for XL184 and placebo arm, respectively) and AE or SAE unrelated to disease progression (16% and 8% for XL184 and placebo arm, respectively).

There were 36% and 41% of the patients discontinued follow-up for XL184 and placebo arm, respectively. The primary reason for discontinuation for follow-up was due to death (28% and 24% for XL184 and placebo arm, respectively).

Table 3 Applicant's Summary of Patient Disposition

Subject Disposition	XL184 n (%)	Placebo n (%)
Total number of randomized subjects (ITT population)	219	111
Subjects still on study treatment at data cut-off date[a]	98 (44.7)	15 (13.5)
Subjects who received study treatment (Safety population)	214 (97.7%)	109 (98.2%)
Subjects who discontinued study treatment	121 (55.3)	96 (86.5)
Disease progression per investigator	58 (26.5)	67 (60.4)
Per mRECIST	51 (23.3)	60 (54.1)
Clinical deterioration	7 (3.2)	7 (6.3)
AE or SAE unrelated to disease progression	35 (16.0)	9 (8.1)
Death	11 (5.0)	5 (4.5)
Subject request other than AE	9 (4.1)	13 (11.7)
Investigator decision other than AE	2 (0.9)	0
Applicant decision other than AE	0	0
Protocol violation	0	0
Lost to follow-up	0	0
Randomized by not treated	5 (2.3)	2 (1.8)
Other		
Subjects who discontinued follow-up	78 (35.6)	45 (40.5)
Primary reason for discontinuation from follow-up		
Death	62 (28.3)	27 (24.3)
Subject withdrew consent	13 (5 9)	13 (11.7)
Lost to follow-up	3 (1.4)	4 (3.6)
Other	0	1 (0.9)

[a] The data cut-off date was 6/15/2011.

The applicant defined the protocol deviations including patients who did not meet the inclusion/exclusion criteria, who received study medication past disease progression, who had incorrect dose and received prohibited study medication. A total of 28 (13%) and 11 (10%) patients in Cabozantinib and placebo arms, respectively had protocol deviation reported.

The distribution of the demographic Characteristics appears to be comparable between treatment arms. Approximately, 67 % of the patients were male, 77% of the patients were younger than 65 years old and 89% of the patients were White. There were more former or current smokers in Cabozantinib arm (49% and 33% in Cabozantinib and placebo arms, respectively). The average weight of the ITT population was 74 kg.

Table 4 Summary of Demographic Characteristics

Subject Characteristic	**XL184 N=219**	**Placebo N=111**
Sex (n, %)		
Male	151 (68.9)	70 (63.1)
Female	68 (31.1)	41 (36.9)
Age (years)		
Mean (standard deviation) Median	54.4 (13.3)	53.8 (13.39)
(range)	55.0 (20-86)	55.0 (21-79)
Age Category (years)		
<65 years	171 (78.1)	83 (74.8)
≥65	48 (21.9)	28 (25.2)
Race (n, %)[a]		
American Indian or Alaska Native	0	0
Asian	9 (4.1)	6 (5.4)
Black or African American	2 (0.9)	1 (0.9)
Native Hawaiian or Other Pacific Islander	0	0
White	196 (89.5)	99 (89.2)
Other	5 (2.3)	1 (0.9)
Not reported	7 (3.2)	4 (3.6)
Smoking (n, %)		
Never	112 (51.1)	74 (66.7)
Former	85 (38.8)	34 (30.6)
Current	22 (10.0)	3 (2.7)
Weight (kg)		
Mean (standard deviation)	72.97 (17.94)	74.77 (19.67)
Median (range)	71.45 (30.4-137.9)	73.20 (41.0-135.9)

[a] Patients can report more than 1 race.

In general, the distribution of baseline Disease Characteristics also appear to be comparable between treatment arms. Approximately 54% of the patients had ECOG performance score 0 and approximately 95% of the patients had measurable disease. There was some imbalance observed in the baseline sum of the longest diameter of the tumor lesions between treatment arms. The Cabozantinib arm appears to have a longer median baseline sum of the longest diameter (106 and 89 mm for the Cabozantinib and placebo arm, respectively).

Table 5 Applicant's Summary of Baseline Disease Characteristics

Characteristic	XL184 N=219	Placebo N=111
ECOG PS (n, %)		
0	123 (56.2)	56 (50.5)
1	86 (39.3)	44 (39.6)
2	9 (4.1)	11 (9.9)
Number of subjects with measurable disease (n, %)		
Investigator	211 (96.3)	107 (96.4)
IRC	208 (95.0)	104 (93.7)
Baseline sum of the longest diameter - IRC (mm)		
N	208	104
Mean (SD)	116.26 (73.756)	103.62 (67.721)
Median (min max)	106.15 (10.7-420.2)	88.80 (10.6-329.5)

ECOG PS=Eastern Cooperative Oncology Group performance status; IRC=Independent Radiology Review Committee;

Reviewer's Comments:

A Cox's proportional hazards model using treatment indicator and a dichotomized variable based on the baseline sum of the longest diameter (< median vs. ≥ median) in the model was fitted to evaluate whether the discrepancy in the median baseline sum of the longest diameter between treatments may affect the primary efficacy results and the results do not appear to change much (HR=0.28, 95% CI=[0.19, 0.39]).

The distribution of baseline MTC disease history, including cancer stage at enrollment and current sites of metastasis disease or number of organs and anatomic locations involvement also appears to be comparable between treatment arms. The most common sites of metastasis were lymph nodes (79%), liver (66%) and lung (55%). Approximately, 95% of the patients had stage IVc at enrollment and 87% of the patients had ≥ 2 organ and anatomic locations involvement The median time since diagnosis of MTC to randomization appears to be longer in the placebo arm as compared with that of the Cabozantinib arm (4.4 vs. 3.6 years, respectively).

Table 6 Applicant's Summary of Baseline MTC Disease History

	XL184 N=219	Placebo N=111
AJCC[a] Stage at enrollment – n (%)		
III	0	1 (0.9)
IVa	4 (1.8)	1(0.9)
IVb	2 (0.9)	1(0.9)
IVc	210 (95.9)	105 (94.6)
Unknown	3 (1.4)	3 (2.7)
Current sites of metastatic disease at enrollment – n (%)		
Bone	219 (100)	110 (99.1)
Lymph nodes	112 (51.1)	56 (50.5)
Cervical	175 (79.9)	86 (77.5)
Mediastinum	111 (50.7)	65 (58.6)
Other	130 (59.4)	60 (54.1)
Liver	58 (26 5)	31 (27.9)
Brain	152 (69.4)	67 (60.4)
Neck	5 (2.3)	2 (1.8)
Lung	37 (16 9)	12 (10.8)
Pelvis	116 (53.0)	64 (57.7)
Other	5 (2.3)	5 (4.5)
Unknown	24 (11.0)	20 (18.0)
	0	0
Number of organs involved at enrollment (n,%)		
0	28 (12.8)	15 (13.5)
≥2	191 (87.2)	96 (86.5)
Years since initial diagnosis of MTC - N	219	111
Mean (SD)	5.9 (6.4)	7.3 (7.9)
Median (min,max)	3.6 (0.1, 33.7)	4.4 (0.2, 48.4)
Years since diagnosis of mMTC – N	218	110
Mean (SD)	3.6 (4.7)	4.6 (5.9)
Median (min,max)	1.9 (0.1, 33.7)	2.0 (0.04, 29.2)

[a] AJCC: American Joint Committee on Cancer

Reviewer's Comments:

A Cox's proportional hazards model using treatment indicator and a dichotomized variable based on the time from diagnosis of MTC to randomization (< median vs. ≥ median) in the model was fitted to evaluate whether the discrepancy in the median time from diagnosis of MTC to randomization between treatments may affect the primary efficacy results and the results do not appear to change much (HR=0.25, 95% CI=[0.17, 0.36]).

Number of patients with prior thyroidectomy, types of prior therapy for MTC, number of prior anti-cancer regimens, number of prior systemic therapy for MTC as well as prior TKI (Tyrosine kinase inhibitor) appear to be comparable between treatment arms. Approximately 92% of the patients had prior thyroidectomy, 58% of the patients did not have prior anticancer therapy reported and 39% had prior systemic therapy for MTC. The majority of patients (65%) received either systematic cancer therapy or radiation or both therapies for MTC. Approximately 21% of the patients received prior TKI (tyrosine kinase inhibitor).

Table 7 Summary of Prior Cancer Therapy

	XL184 N=219	Placebo N=111
Number of subjects (%) with prior thyroidectomy	201 (91.8)	104 (93.7)
Number of subjects (%) with prior therapy for MTC		
Prior anticancer therapy only[a]	37 (16.9)	23 (20.7)
Prior radiation therapy only	56 (25.6)	27 (24.3)
Prior anticancer therapy and radiation[a]	48 (21.9)	25 (22.5)
No prior therapy reported	78 (35.6)	36 (32.4)
# of Prior anticancer therapy regimens[b]		
0	128 (58.5)	62 (55.9)
1	36(16.4)	18 (16.2)
≥2	55(25.1)	31 (27.9)
Number of subjects (%) with prior systemic therapy for MTC	81 (37.0)	47 (42.3)
Prior tyrosine kinase inhibitor status (n, %)[c]		
Yes	44 (20.1)	24 (21.6)
No	171 (78.1)	86 (77.5)
Unknown	4 (1.8)	1 (0.9)

[a] Prior anticancer therapy includes systemic treatment and chemembolization but not radiation therapy.
[b] # of prior anticancer regimens for all cancer (i.e. not only for MTC).
[c] Based on information collected on the Prior Tyrosine Kinase Inhibitor Exposure CRF page.

Only 61% of the patients had RET mutation status ascertained. Approximately 46% and 52% of the patients were RET mutation positive for Cabozantinib arm and placebo arm, respectively. The majority of the MTC disease type was sporadic (87% and 85% for Cabozantinib and placebo arm, respectively). Also, 34% and 39% of the patients for Cabozantinib and placebo arm, respectively, had positive RET M918T mutation status.

Table 8 Summary of RET Genotyping Results

	XL184 N=219 n(%)	Placebo N=111 n(%)
RET Mutation Status[a]		
Positive	101 (46.1)	58 (52.3)
Negative	31 (14.2)	10 (9.0)
Unknown	87 (39.7)	43 (38.7)
MTC Disease Type[b]		
Hereditary	12 (5.5)	8 (7.2)
Sporadic	191 (87.2)	94 (84.7)
Unknown	16 (7.3)	9 (8.1)
RET M918T Mutation Status[c]		
Positive	75 (34.2)	43 (38.7)
Negative	67 (30.6)	30 (27.0)
Unknown	77 (35.2)	38 (34.2)

a *RET* Mutation Positive: evidence of *RET* mutation in either blood or tumor sample. *RET* Mutation Negative: adequate sequence of tumor sample without evidence of *RET* mutation.
b Hereditary: evidence of *RET* mutation in blood DNA sample. Sporadic: adequate sequence of blood or tumor DNA sample with no evidence of *RET* mutation.
c M918T Mutation Positive: presence of a *RET* M918T mutation in either blood or tumor DNA sample. M918T Mutation Negative: adequate *RET* exon 16 sequence data from tumor DNA sample with no evidence of M918T mutation.

3.2.3 Statistical Methodologies

Analysis Population

The following statistical analysis populations were proposed for the efficacy analyses:

- **Intent-to-treat**: Contains all patients who were randomized regardless of whether any study treatment or the correct study treatment is administered.
- **Safety**: Contains all patients who receive any amount of treatment. The safety population will be analyzed according to the actual treatment received.
- **Per-protocol**: Contains all patients in the Safety population and who met the criteria as defined on page 11 of 41 of Section 16.1.9 Documentation of Statistical Methods.

Primary efficacy Analysis

The statistical analysis method for PFS was based on the stratified log rank test at a 2-sided 0.05 α level using the stratification factors documented based on the IVRS. The primary analysis of PFS would be conducted after at least 138 events had been observed. The median duration of PFS and the associated 95% confidence interval for each treatment arm would be estimated using the Kaplan-Meier method. The HR would be estimated based on the stratified Cox's proportional hazards model.

Sensitivity analyses for the Progression Free Survival

A list of sensitivity analyses for PFS based on different definitions of progression events and censoring rules are shown in the table shown below (Table 9).

PFS2 analysis defines the date of progression (defined by IRC) as the scheduled tumor assessment (or the next scheduled tumor assessment date if between assessments) instead of the recorded date of progression. This analysis was used to correct for potential ascertainment bias in follow-up schedule between treatments.

PFS3 analysis was performed to assess the investigator assessment of radiographic progression. This analysis did not consider the clinical progression events.

PFS4 analysis was performed to assess PFS based on the investigators' claim. The analysis is similar to PFS3 analysis except that the clinical deterioration and the initiation of subsequent events were counted as progression events.

PFS5 analysis was performed by censoring patients at the last tumor assessments dates (i.e. ignore the censoring scheme as used in the primary analysis).

Table 9 Summary of the Sensitivity Analyses for Progression Free Survival

Sensitivity analyses [a]	Endpoint	Description of the analysis
Primary (PFS1)	IRC-assessed PFS	See the description in the texts.
PFS2	IRC-assessed PFS	Events defined based on the date of progression (by IRC) as the scheduled tumor assessment (or the next scheduled tumor assessment date if between assessments) rather than the recorded date of progression, i.e. All PD events were moved to a multiple of 12 weeks.
PFS3	Investigator-assessed PFS (radiological assessment)	Events determined by the investigator assessment of radiographic progression. This analysis did not consider the clinical progression events.
PFS4	Investigator-assessed PFS (include clinical deterioration)	Events defined based on the investigators' claim (i.e. similar to PFS3 except that the clinical deterioration and the initiation of subsequent events were counted as progression events.)
PFS5	IRC-assessed PFS	Censored patients at the last tumor assessment dates prior to data cutoff, i.e. ignore the censoring reasons as indicated for the primary PFS analysis.
Per-protocol	IRC-assessed PFS	Similar to the primary PFS analysis except that this analysis is based on per-protocol population

[a] PFS2-PFS4 correspond to the applicant's sensitivity analysis plan for PFS described in the SAP.

Secondary efficacy Analysis

Objective response rate (ORR) and overall survival (OS) are the two key secondary endpoints. If the primary efficacy analysis shows statistical significance, the two key secondary analyses would then be tested in parallel. The ORR and OS will be tested at the 0.01 and 0.04 levels, respectively. An interim analysis of OS was described in the Interim Analysis section of Section 3.2.1 Study Design and Endpoints in this review.

The statistical analysis method for OS was also based on the stratified log rank test and would be performed based on a 2-sided 0.04 level. The primary analysis of OS would be conducted after at least 217 deaths had been observed. The median duration of OS and the associated 95% and 96% confidence interval for each treatment arm would be estimated using the Kaplan-Meier method. The HR would be estimated based on the stratified Cox's proportional hazards model.

Formal hypothesis testing for ORR between treatment arms was performed using Chi-squared test at a 2-sided significance level of 0.01. If a sufficient number of responders had been observed, Cochran-Mantel-Haenszel method would be used for the analysis of ORR. Point estimates of ORR, the difference in response rates between treatment arms and associated 95% and 99% confidence intervals for the response rate and the difference would be summarized.

3.2.4 Results and Conclusions

The efficacy results for study XL184-301 will be described in this section.

3.2.4.1 Efficacy Endpoint Analyses

Progression Free Survival

The primary efficacy analysis was planned to be performed at the time when 138 PFS events were reached (on 4/6/2011 cut-off date). At the time of the primary PFS analysis, there were 36% and 54% of the PFS events for the Cabozantinib and placebo arms, respectively. The estimated hazard ratio for PFS was 0.28 (95% CI=0.19, 0.40) in favor of the Cabozantinib arm. The Cabozantinib arm appears to have longer median duration of PFS as compared with placebo arm (11.2 months vs. 4 months, respectively).

Table 10 Reviewer's Summary of Progression Free Survival (based on the 139th events; 4/6/2011 cutoff date)

	XL184 N=219	Placebo N=111
Number (%) of Subjects		
Censored	140 (63.9)	51 (46.0)
Event	79 (36.1)	60 (54.1)
Death	21 (9.6)	10 (9.0)
Progressive disease	58 (26.5)	50 (45.0)
Duration of progression free survival (mon.) Median (95% CI) a	11.2 (8.4,13.7)	4.0 (3.0, 5.4)
Range	0.0+ - 22.1	0.0+ - 16.7
p-value (stratified log-rank test)b	<0.0001	
Hazard ratio (95% CI; stratified)c	0.28(0.19, 0.40)	

Note: 139 events occurred by the date of the 138th event.
+ indicates a censored observation; CI=confidence interval; IRC=Independent Radiology Review Committee.
a Median and percentiles are based on Kaplan-Meier survival estimates.
b Stratification factors include age at randomization (≤65, >65) and prior tyrosine kinase inhibitor status (yes, no).
c Estimated using the Cox proportional hazard model adjusted for stratification factors.

Reviewer's comments:

- *There was no differential censoring distribution between treatment arms for PFS based on a Cox's proportional hazards model including treatment indicator in the model and using a reversed censoring indicator (hazard ratio=0.84, 95% CI=0.59, 1.20).*
- *There were a total of 184 PD events or deaths based on the 6/15/2011 cutoff date. Based on the statistical plan, the patients were censored for PFS among patients who took anti-cancer therapy, died after missing more than 1 tumor assessments, had PD after missing more than 1 tumor assessments. The following table summarizes the distributionof patients by censoring reasons. After these 25 patients were censored based on reasons indicated in the table, there were 159 PFS events at the data cutoff date (6/15/2011). The primary PFS analysis was based on the first 139 patients among these PFS events (dated 4/6/2011).*

Table 11 Reviewer's Summary of Censoring Distribution by censoring reasons (based on 6/15/2011 cutoff date)

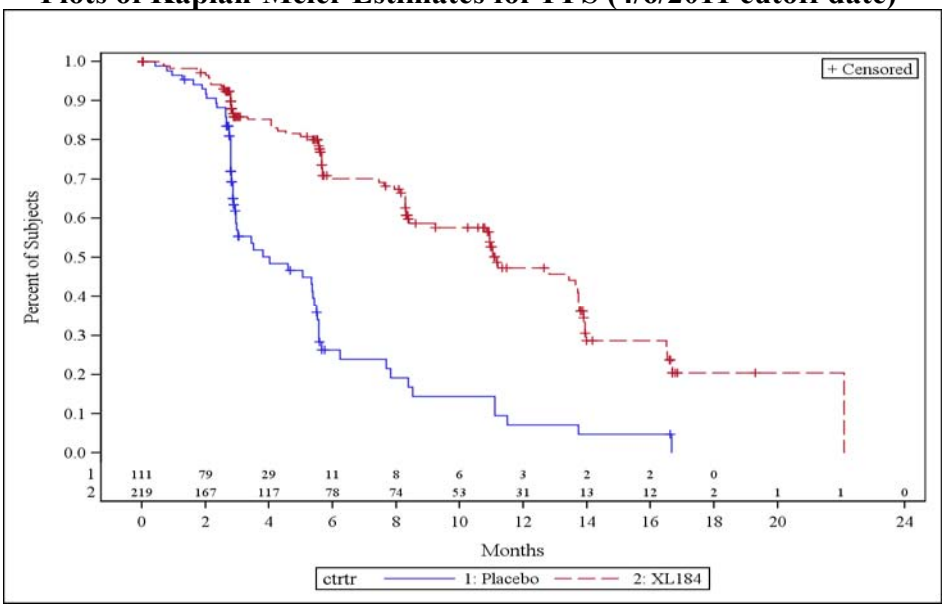
Censoring reasons	XL184 (n=219)	Placebo (n=111)	Total (n=330)
# Patients who had either PD or died	109 (49.8%)	75 (67.6%)	184 (55.8%)
#patients who took anti-cancer therapy	13 (5.9%)	8 (7.2%)	21 (6.4%)
#patients who died after missing >1 assessment periods	3 (1.4%)	0	3 (1%)
#Patients who had PD after 2 missing assessment time	0	0	0
# patient who had missing >=2 assessment after randomization [a]	1 (<1%)	0	1 (<1%)

[a] These patients did not have post baseline tumor assessments.

- *There were 35 patients in the study whose disease progression status was determined by the investigator per amendment 2 (rather than determined by IRC prior to amendment 1). Among these 35 patients, two patients from Cabozantinib arm had PFS events because of deaths. There are a total of 3 deaths in these 35 patients, all from the Cabozantinib arm. One patient was not counted as having a PFS event in the primary analysis because his death date passed 4/6/11 cutoff date for the primary PFS analysis. To evaluate the impact of these patients on the primary PFS analysis results, this reviewer ran a Cox's PH model (similar to the primary analysis for PFS) including amendment number (1 vs 2) as an additional stratification factor, the results are still consistent with the primary analysis results (HR=0.26, 95% CI=0.181, 0.381).*

The corresponding plots for the Kaplan-Meier estimates are presented in the following figure :

Figure 1 Plots of Kaplan-Meier Estimates for PFS (4/6/2011 cutoff date)



An exploratory analysis of PFS based on the database cutoff date (6/15/2011) was also performed. The hazard ratio estimate was very similar to the result from the primary efficacy analysis (HR=0.28; 95% C.I.=0.19, 0.39).

Table 12 Reviewer's Summary of Progression Free Survival (based on the 159th events; 6/15/2011 cutoff date)

	XL184 N=219	Placebo N=111
Number (%) of Subjects		
Censored	127 (58.0)	44 (39.6)
Event	92 (42.0)	67 (60.4)
Duration of progression free survival (weeks)		
Median (95% CI) [a]	12.4 (10.8, 13.7)	4.0 (3.0, 5.5)
Range	0.0+ - 22.1+	0.0+ - 19.4
p-value (stratified log-rank test) [b]	<0.0001	
Hazard ratio (95% CI; stratified) [c]	0.28(0.19, 0.39)	

+ indicates a censored observation; CI=confidence interval; IRC=Independent Radiology Review Committee.
a Median and percentiles are based on Kaplan-Meier survival estimates.
b Stratification factors include age at randomization (≤65, >65) and prior tyrosine kinase inhibitor status (yes, no).
c Estimated using the Cox proportional hazard model adjusted for stratification factors.

Evaluation of Concordance and Discordance of the IRC and Investigator Assessments

Based on the 6/15/2011 cutoff date, the percentage of patients had PD or non-PD determined by both the IRC and investigators (concordance) are 78.5% (see the table below).

Table 13 Reviewer's Summary of Concordance/Discordance in Progressive Disease Status (based on 6/15/2011 cutoff date)

Status	If progressed	XL184 (n=219) n(%)	Placebo (n=111) n(%)	Total (n=330) n(%)
Concordance	Progressive Disease	44 (20.1)	48 (43.2)	92 (27.9)
	Not Progressive Disease	131(59.8)	36(32.4)	167 (50.6)
Discordance	IRC progressed/INV not progressed	25 (11.4)	8 (7.2)	33 (10.0)
	INV progressed/IRC not progressed	19 (8.7)	19 (17.1)	38 (11.5)

Among these patients with concordant PD status, 75% had concordant timing of the PD. Among patients with discordant PD status (n=23, 25%), 20 patients had IRC determined timing of PD earlier than those determined by the investigators and such trend appears to be similar between treatment arms.

Table 14 Reviewer's Summary of Concordance/Discordance in Timing of the Progressive Disease Status (PD) Among Patients who had PD Determined by Both IRC and Investigators (based on 6/15/2011 cutoff date)

PD determined by both IRC and INV	XL184 (n=219) n(%)	Placebo (n=111) n(%)	Total (n=330) n(%)
Progressive Disease	44 (20)	48 (43)	92 (28)
Concordance in PD time	31 (14)	38 (34)	69 (21)
Discordance in PD time	13 (6)	10 (9)	23 (7)
IRC earlier than INV	11 (5)	9 (8)	20 (6)
INV earlier than IRC	2 (1)	1 (1)	3 (1)

Sensitivity Analyses For PFS

The sensitivity analysis results for PFS are summarized in the following table. All results, based on two different cut-off date, appear to be supportive of the primary PFS analysis (HR=0.28). The magnitude of the effect based on the difference in median PFS duration ranged from 5.7 to 10.8 months also appears to be supportive of the primary result (difference in median PFS time =7.2 months; with majority of the PFS median time difference > 7.2 months). The result based on the investigator claims (i.e. PFS4, including clinical deterioration and treating patients who took anti-cancer therapy as a PFS event) appears to have the largest hazard ratio estimate.

Table 15 Reviewer's Sensitivity Analyses for PFS

Cut-off Date	Sensitivity # [a]	#event/total XL184 : Plc	Hazard Ratio	Median (95% CI) XL184	Median (95% CI) Placebo	Difference (XL184-Plc)
4/6/11	Primary (IRC)	79/219:60/111	0.27(0.19,0.40)	11.2(8.4,13.7)	4.0(3.0,5.4)	7.2
	PFS2 (IRC)	79/219:60/111	0.28(0.20,0.41)	11.1(10.9,13.7)	5.4(2.9,5.6)	5.7
	PFS3 (INV)	70/219:66/111	0.29(0.20,0.42)	13.8(10.7,16.3)	3.1(2.9,5.4)	10.7
	PFS4 (INV)- w/ clinical events	84/219:80/111	0.31(0.23,0.43)	11.2(8.3,13.9)	3.0(2.8,4.3)	8.2
	PFS5 (IRC)	93/219:66/111	0.29(0.20,0.41)	12.8(10.8,13.9)	5.3(3.0,5.6)	7.5
	Per-protocol	76/198:59/101	0.26(0.18,0.38)	11.2(9.2,13.7)	3.8(2.9,5.4)	7.4
6/15/2011	Primary (IRC)	92/219:67/111	0.27(0.19,0.39)	12.4(10.8,13.7)	4.0(3.0,5.5)	8.4
	PFS2 (IRC)	92/219:67/111	0.28(0.20,0.40)	12.4(11.0,13.8)	5.4(3.0,5.6)	7.0
	PFS3 (INV)	84/219:75/111	0.28(0.20,0.39)	13.8(11.0,14.9)	3.0(2.9,5.4)	10.8
	PFS4 (INV)- w/ clinical events	100/219:91/111	0.29(0.22,0.40)	12.4(8.9,13.9)	3.0(2.8,3.5)	9.5
	PFS5 (IRC)	109/219:75/111	0.29(0.21,0.41)	11.2(8.9,13.7)	5.1(3.0,5.5)	6.1
	Per-protocol	88/198:66/101	0.26(0.19,0.38)	12.8(10.9,13.8)	4.0(2.9,5.4)	8.8

[a] PFS2-PFS4 correspond to the applicant's sensitivity analysis plan for PFS described in the SAP.

Overall Survival

There were 96 deaths (i.e. 44% information level) at the clinical cutoff date for the PFS analysis (dated 6/15/2011). The median follow-up time in the ITT population was 8.4 and 7.8 months for Cabozantinib and placebo arm respectively. At the cutoff date, the hazard ratio was close to 1 (HR=0.997, 95% CI=0.64, 1.54) which indicates that the overall survival time were similar between treatment arms. The median survival time was 21 months (95% CI=16.6, 28.5) for the Cabozantinib arm while the median survival time for the placebo arm had not yet reached. Based on the Lan-DeMets O'Brien-Fleming alpha spending function,

the significance level for the interim analysis for OS was 0.0009. So, the OS result did not cross the boundary for claiming of the overall survival benefit.

Table 16 Reviewer's Summary of Overall Survival (based on 6/15/2011 cutoff date)

	XL184 N=219	Placebo N=111
Number (%) of Subjects		
Censored Death	153 (69.9) 66 (30.1)	81 (73.0) 30 (27.0)
Duration of overall survival (months)		
Median (95% CI) [a] Min- Max	21.1 (16.59, 28.52) 0.0+ - 29.5+	NA (14.32, NA) 0.1+-32.1+
p-value (stratified log-rank test)[b]	0.989	
Hazard ratio (95% CI; stratified)[c]	0.997 (0.644, 1.542)	

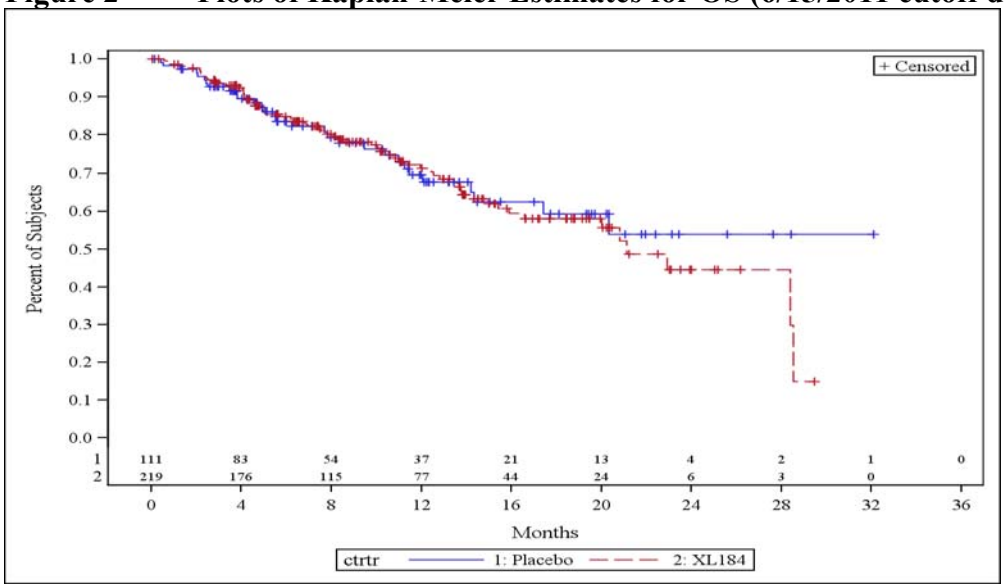
+ indicates a censored observation; CI=confidence interval;
a Median and percentiles are based on Kaplan-Meier survival estimates.
b Stratification factors include age at randomization (≤65, >65) and prior tyrosine kinase inhibitor status (yes, no).
c Estimated using the Cox proportional hazard model adjusted for stratification factors.

Reviewer's Comments:

- *There was no differential censoring distribution between treatment arms for overall survival based on the Cox's proportional hazards model including treatment indicator in the model and using a reversed censoring indicator (hazard ratio=0.90, 95% CI=0.68, 1.18).*

The Plots for the Kaplan-Meier estimates are presented below which show two survival curves overlapped with each other. The Cabozantinib arm appears to have more deaths occurred as compared with the placebo arm toward the tail of the curves. However, the results toward the end of the K-M curve may not accurately estimate due to the limited number of events.

Figure 2 Plots of Kaplan-Meier Estimates for OS (6/15/2011 cutoff date)



Reviewer's Comments:

- *Based on the 120-day updated data (with 66 more deaths, i.e. 75% of the information level), the hazard ratio estimate for the OS was 0.825 (95% CI=0.598, 1.140) and the nominal p-value was 0.2993 based on the stratified log rank test. The median OS time was 26.02 (95% CI=22.90, 30.72) and 20.34 (95% CI=16.39, 26.68) for the Cabozantinib arm and placebo arm, respectively.*

Table 17 Reviewer's Summary of Overall Survival (based on 120-day updated data with 6/15/2012 cutoff date)

	XL184 N=219	**Placebo N=111**
Number (%) of Subjects		
Death	103 (47)	59 (53.0)
Duration of overall survival (months)		
Median (95% CI) [a]	26.02 (22.90, 30.72)	20.34 (16.39, 26.68)
p-value (stratified log-rank test)[b]	0.2432	
Hazard ratio (95% CI; stratified)[c]	0.825 (0.598, 1.14)	

CI=confidence interval;
a Median and percentiles are based on Kaplan-Meier survival estimates.
b Stratification factors include age at randomization (≤65, >65) and prior tyrosine kinase inhibitor status (yes, no).
c Estimated using the Cox proportional hazard model adjusted for stratification factors.

Best Overall Response Rate

The proportion of IRC-determined measurable disease at baseline was comparable between treatment arms (95% vs. 94% for the Cabozantanib arm and placebo arm, respectively). There were more patients in placebo arm to have disease progression as compared with the Cabozantanib arm (32% vs. 8%, respectively). The ORRs were 26.5 % and 0% for the Cabozantanib arm and placebo arm, respectively. The p-value based on the stratified Cochran-Mantel-Haenszel test was significant ($p<0.0001$).

The disease stabilization rates were 53% for the Cabozantanib arm and 13% for the placebo arm, which was also in favor of the Cabozantanib arm.

Table 18 Reviewer's Summary of Objective Response Rate

Subjects in ITT Population	**XL184 N=219**	**Placebo N=111**
Best Overall Response (n, %)[a]		
Confirmed complete response (CR)	0	0
Confirmed partial response (PR)	58 (26.5)	0
Stable disease (SD)	100 (45.7)	52 (46.9)
Progssionsive disease	18 (8.2)	35 (31.5)
Unable to evaluate	5 (2.3)	1 (0.9)
Missing[b]	38 (17.4)	23 (20.7)
Objective Response Rate (ORR=CR+PR)		
n (%)	58 (26.5)	0
95% confidence interval	20.8%, 32.9%	NA
99% confidence interval	19.2%, 34.9%	NA
p-value (stratified Cochran-Mantel-Haenszel test)[c]	<0.0001	
Duration of Response(month)	14.7 (11.1, 19.3)	NA
Disease Stabilization Rate (DSR=ORR+SD)		
n (%)	115 (52.5)	14 (12.6)
95% CI	45.7%, 59.3%	7.1%, 20.3%
p-value (stratified Cochran-Mantel-Haenszel test) [c]	<0.0001	

IRC=Independent Radiology Review Committee; NA=not available
a Best overall response determined by IRC using mRECIST criteria.
b Missing=no qualifying post-baseline assessment for overall response.
c Stratification factors : age and prior tyrosine kinase inhibitor status.

Reviewer's Comments:

- *The denominators used to derive the rate measures in this table are based on the ITT population.*

Biomarker Analyses

Exploratory analyses based on tumor marker data : CEA (Carcinoembryonic antigen), Calcitonin, were performed. It is noted that there are elevated calcitonin and CEA levels in

patients with medullary thyroid cancer. This section will evaluate if the treatment has an effect on the reduction of the calcitonin and CEA levels .

There were 78% and 64% of the patients with available baseline and week 12 CEA values. Among these available CEA data at week 12, the Cabozantinib arm had 38 % decrease from baseline and the placebo arm had 38% increase from baseline. Similarly for the calcitonin level, with over 35% missing data, there were 60% decrease and 23% increase from baseline for the Cabozantinib and placebo arm, respectively. Both results show nominally significant difference between treatment arms.

Table 19 Reviewer's Summary of Tumor Marker Data, Change from Baseline at Week 12

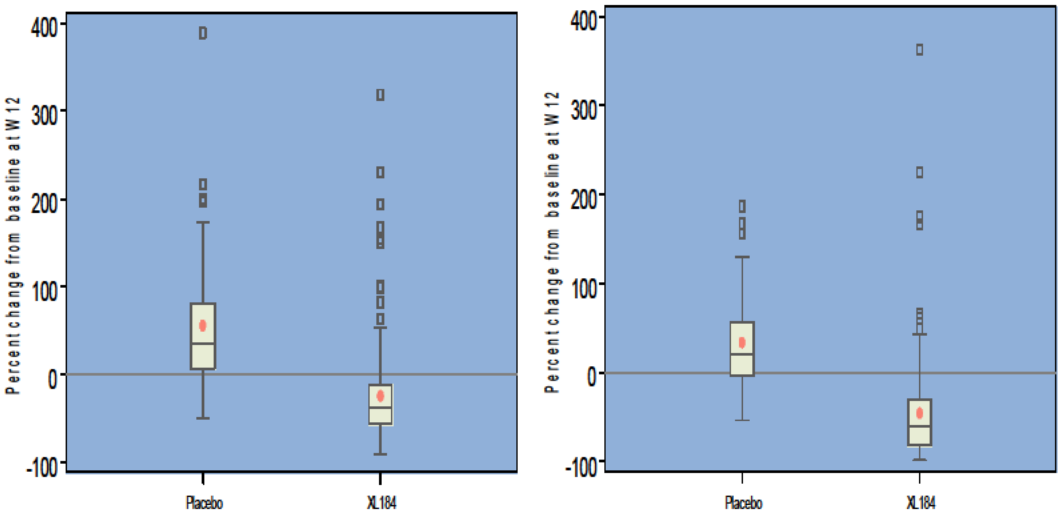
	XL184 **N=219** **Median (q1, q3)**	**Placebo** **N=111** **Median (q1, q3)**
CEA µg/L[n (%)] Baseline W12 Change from baseline Percent Change from baseline	170 (78%) 120.7 (33.5,422.7) 56.4 (21.4, 260.9) -23.7 (-143.1, -3.2) -38.0 (-56.1, -11,5)	71 (64%) 153.1 (32.3,478.2) 221.8 (69.5, 962.7) 35.6 (4.1, 269.6) 38.0 (8.9, 104.0)
p-value[a]	<0.0001	
Calcitonin pmol/L[n (%)] Baseline W12 Change from baseline Percent Change from baseline	140 (64%) 2298.1 (544.5,5754.0) 584.8 (177.3, 2671.5) -1188 (-3071.0,-135.4) -60.2 (-81.7, -29.5)	61 (55%) 3886.0(792.0,9237.4) 4968.0 (1219.0,11716.0) 322 (-0.5, 3941.3) 22.7 (-2.3, 67.3)
p-value[a]	<0.0001	

[a]Based on Wilcoxon Rank Sum Test ; nominal p-value from exploratory subgroup analyses.

The box-plots shown in the following figures also demonstrated greater percent reduction from baseline in the Cabozantinib arm as compared with that from the placebo arm at week 12.

Figure 3 Percent Change from Baseline in CEA and Calcitonin level at Week 12

CEA **Calcitonin**



Reviewer's comments:

Three patients who had extreme values of the change from baseline of CEA/calcitonin level were not plotted in the box plot for a better presentation of the graph. The corresponding percent change from baseline of the CEA levels for these patients are 1257.58, 677.40 and 530.16 (all from the placebo arm). Similarly, the corresponding percent change from baseline of the calcitonin levels for these patients are 537.651, 433.913 and 535.922 (all from the placebo arm).

3.3 Evaluation of Safety

The safety evaluation was not performed in this statistical review. Please refer to the clinical review for more details for the safety assessments.

3.4 Benefit-Risk Assessment

The benefit-risk assessment was not performed in this statistical review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Subgroup analyses based on gender, race, age group and geographic regions are performed in this section.

4.1 Gender, Race, Age, and Geographic Region

Gender

The hazards ratio estimates based on PFS for both male and female subgroups were small with the upper bound of the 95% CIs being less than 1 which appears to support the favorable treatment effect in the Cabozantinib arm for both gender subgroups.

Table 20 Reviewer's Summary of Hazard Ratios for PFS by Gender

		XL184 N=219	Placebo N=111
Male	Number of events / total	60/151	37/70
	HR (95% CI) [a]	0.33(0.21,0.51)	
Female	Number of events / total	19/68	23/41
	HR (95% CI) [a]	0.22(0.12,0.43)	

[a]From Cox's Proportional Hazards Model (unstratified)

Race

The hazard ratio estimate based on PFS from the White subgroup was smaller than 1 which also demonstrate favorable treatment effect in the Cabozantinib arm. The hazard ratio estimate based on the non-White subgroup appears to be larger than that in the White subgroup. However, the interpretation of the non-White subgroup should be taken with caution because only 10% of the patients were in this subgroup and the 95% CIs are wide.

Table 21 Reviewer's Summary of Hazard Ratios for PFS by Race

		XL184 N=219	Placebo N=111
White	Number of events / total	70/196	55/99
	HR (95% CI) [a]	0.28(0.19,0.40)	
Non-White	Number of events / total	9/23	5/12
	HR (95% CI) [a]	0.59(0.19,1.84)	

[a]From Cox's Proportional Hazards Model (unstratified)

Age

Similarly, the hazard ratio estimates based on PFS for both age subgroups were less than 1 with the upper bound of the 95% CIs being smaller than 1 which appears to support the treatment benefit in the Cabozantinib treated arm for both younger and older patient subgroups.

Table 22 Reviewer's Summary of Hazard Ratios for PFS by Age Subgroup

		XL184 N=219	Placebo N=111
<65 years old	Number of events / total	59/171	44/83
	HR (95% CI) [a]	0.29(0.19,0.44)	
≥65 years old	Number of events / total	20/48	16/28
	HR (95% CI) [a]	0.33(0.16,0.66)	

[a]From Cox's Proportional Hazards Model (unstratified)

Geographic Region

The hazard ratio estimates based on PFS for Europe and North American regions were both less than 1 with the upper bound of the 95% CIs being smaller than 1. These results appear to support the treatment benefit in the Cabozantinib arm in these regions. The hazard ratio estimate for the rest of world was also small, however, only 13% of the population were in this subgroup, the 95% CI was wide and the interpretation of this subgroup should be taken with caution.

Table 23 Reviewer's Summary of PFS results by Geographic Regions

		XL184 N=219	Placebo N=111
Europe	Number of events / total	48/124	37/60
	HR (95% CI) [a]	0.26(0.16,0.42)	
North America	Number of events / total	26/69	15/33
	HR (95% CI) [a]	0.32(0.16,0.63)	
Rest of the World	Number of events / total	5/26	8/18
	HR (95% CI) [a]	0.33(0.11,1.03)	

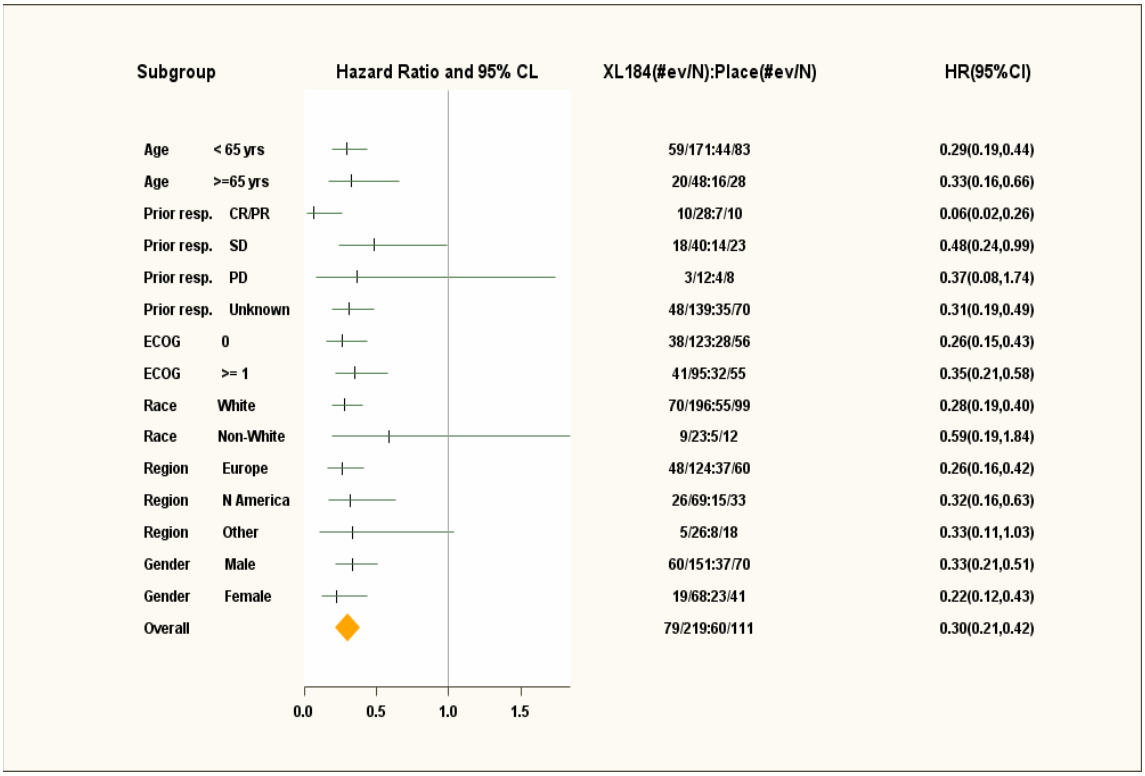
[a]From Cox's Proportional Hazards Model (unstratified)

4.2 Other Special/Subgroup Populations

Forest plots of the hazard ratio estimates based on PFS and the corresponding 95% are shown in this section by demographic information, baseline characteristics and RET mutation status.

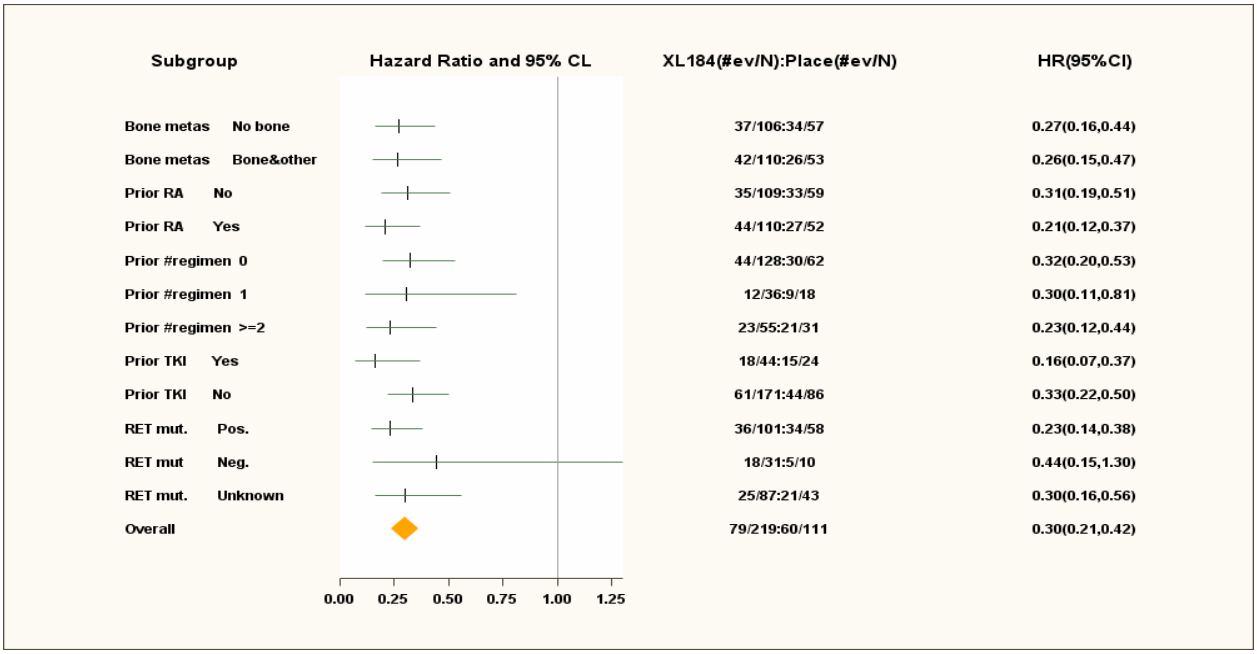
For the forest plots by the demographic information, all the hazard ratio estimates were smaller than 1. Only the hazard ratio estimates for those patients who had prior disease progression status and who were non-White had the upper bound of the 95% cross the reference line 1. However, due to smaller sample size in these subgroups, the 95% CIs are wide and the interpretation should be taken with caution.

Figure 4 Forest Plots based on Hazard Ratio Estimates for PFS by Demographic Information



Similarly, the hazard ratio estimates were all smaller than 1 from the forest plots based on PFS by baseline characteristics. Only the hazard ratio estimate for patients who were RET mutation negative had the upper bound of the 95% CI cross the reference line 1. However, due to smaller sample size in this subgroup, the 95% CIs are wide and one needs to exercise extra caution in the interpretation in the RET mutation negative subgroup.

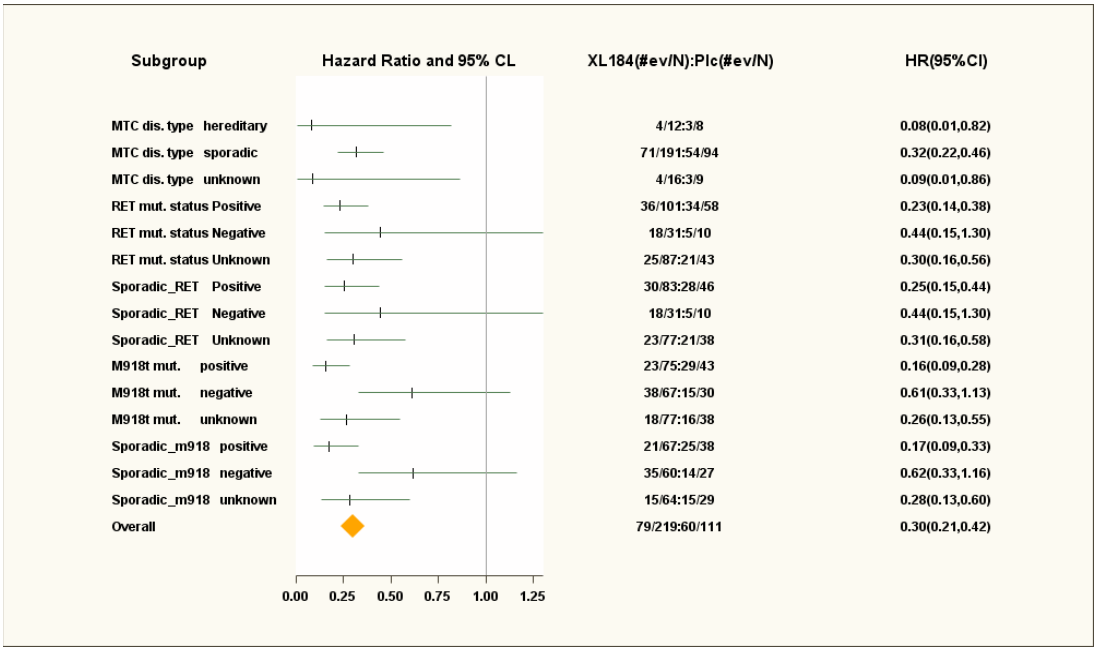
Figure 2 Forest Plots based on Hazard Ratio Estimates for PFS by Baseline Characteristics



To evaluate the consistency of the results in the RET mutation subgroup, subgroup analyses by the RET mutation subgroups were performed. A summary of the RET mutation subgroup was shown in Section 3.2.2 Patient Disposition, Demographic and Baseline Characteristics. Additional subgroup analyses were performed based on RET mutation status using the patient subpopulation with sporadic MTC disease type.

The forest plots based on the subgroup analysis of PFS by RET mutation subgroup is shown in the following figure. All of the hazard ratio estimates and the majority of the upper bound of the 95% confidence interval (except the RET mutation negative, the M918t mutation negative as well as the corresponding mutation among patients with sporadic disease) are less than 1 which indicates the beneficial effect of the Cabozantinib. It is noted that for the hazard ratio estimates from patients with RET mutation negative, M918t mutation negative as well as the corresponding mutation among patients with sporadic disease were all less than 1, but the corresponding 95% CIs were wide due to the smaller sample size in these subgroups.

Figure 2 Forest Plots based on Hazard Ratio Estimates for PFS by RET Mutation Subgroup



5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Based on study XL184-301, the results show significant improvement of the progression free survival. The median PFS time was improved from 4 months (95% CI=3.0, 5.4) in the placebo arm to 11.2 months (95% CI=8.4, 13.7) in the Cabozantinib arm. The hazard ratio estimate was 0.28 (95% CI=0.19, 0.40) in favor of the Cabozantinib arm. The favorable results from the Cabozantinib arm were robust based on various sensitivity analyses and consistent results across different subgroups including RET mutation status. The result based on the objective response rate also demonstrated statistical significance in favor of the Cabozantinib arm (ORR=27% vs. 0% for Cabozantinib arm vs. placebo, respectively). However, based on 44% information level, the result did not demonstrate beneficial effect for the Cabozantinib treated arm based on overall survival (HR=0.997, 95%=0.64, 1.54). Based on updated OS analysis (i.e. 75% information level; 6/15/2012 cutoff date), the result still did not show beneficial effect from the Cabozantinib treated arm (HR=0.825, 95% CI=0.598, 1.140). Nevertheless, the updated OS result appears to

show a positive trend based on the median OS time in favor of the Cabozantinib treated arm (the median OS time was 26.02 [95% CI=22.90, 30.72] and 20.34 [95% CI=16.39, 26.68] for the Cabozantinib arm and placebo arm, respectively). A summary of these primary results are shown below:

Table 24 Reviewer's Summary of PFS, OS and ORR results

Endpoint		XL184 N=219	Placebo N=111
PFS (based on 4/6/2011 cutoff date)	Number (%) of events Progressive disease	79(36)	60 (54)
	Duration of progression free survival (mon.) Median (95% CI) a	11.2 (8.4,13.7)	4.0 (3.0, 5.4)
	p-value (stratified log-rank test)b	<0.0001	
	Hazard ratio (95% CI; stratified)c	0.28(0.19, 0.40)	
OS (based on 6/15/2011 cutoff date)	Number (%) of events death	66(30)	30 (27)
	Duration of progression free survival (mon.) Median (95% CI) a	21.1 (16.6,28.5)	NA (14.3, NA)
	p-value (stratified log-rank test)b	0.989	
	Hazard ratio (95% CI; stratified)c	0.997(0.644, 1.542)	
120-day updated OS (based on 6/15/2012 cutoff date)	Number (%) of events death	103(47)	59 (53%)
	Duration of progression free survival (mon.) Median (95% CI) a	26.0 (22.9,30.7)	20.3 (16.4,26.7)
	p-value (stratified log-rank test)b	0.2434	
	Hazard ratio (95% CI; stratified)c	0.825 (0.598, 1.140)	
ORR	Objective response rate 95% CI	58 (26.5%) (20.8%, 32.9%)	0 NA
	Duration of response(month) 95% CI	14.7 (11.1, 19.3)	NA NA
	p-value (Cochran-MantelHaenszel test) b	<0.0001	

CI=confidence interval;
a Median and percentiles are based on Kaplan-Meier survival estimates.
b Stratification factors include age at randomization (≤65, >65) and prior tyrosine kinase inhibitor status (yes, no).
c Estimated using the Cox proportional hazard model adjusted for stratification factors.

Based on exploratory biomarker analyses, the Cabozantinib treated arm appears to have nominally significant reduction of the CEA and calcitonin levels as compared with the placebo arm.

The main issue from this study is that the OS analysis was not nominally statistically significant. In the protocol, the assumed median OS time in the placebo arm (i.e. 21 months) appears to be

close to the actual result observed based on the updated OS analysis (20.3 months). However, the assumed median OS time in the Cabozantinib arm of the study design was too optimistic (i.e. 33 months for the assumed median OS time vs. 26 months for the actual result observed from the study). The reason of lacking survival benefit is unclear.

5.2 Conclusions and Recommendations

In summary, based on study XL184-301, the results demonstrated statistically significant improvement on progression free survival for the Cabozantinib treated arm in patients with progressive, unresectable locally advanced or metastatic medullary thyroid cancer (MTC). The results appear to be robust based on sensitivity analyses and consistent across many subgroups including RET mutation status. The result based on the objective response rate also demonstrated beneficial effect in favor of the Cabozantinib treated arm. However, the overall survival result does not appear to demonstrate treatment benefit for the Cabozantinib treated arm based on 44% or 75% (120-day update) information level, even though the 120-day updated results appear to show a positive trend in favor of the Cabozantinib treated arm.

In conclusion, this statistical reviewer confirms the applicant's results submitted. Whether the results demonstrate an overall favorable benefit to risk ratio in supporting an indication of the Cabozantinib treatment in patients with progressive, unresectable locally advanced or metastatic medullary thyroid cancer will defer to the clinical review team.

5.3 Labeling Recommendations

This statistical review supported the inclusion of results from the progression free survival and objective response rate, but not the overall survival results, based on the pre-specified statistical analysis plan. The overall survival data is not mature at this time, the current OS data is not reliable. When the pre-specified number of events is reached, OS data should be re-analyzed.

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

/s/

YUAN L SHEN
10/19/2012

KUN HE
10/19/2012
Accepted as a complete review.

RAJESHWARI SRIDHARA
10/19/2012

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: **Applicant: Exelixis** **Stamp Date: 5/21/2012**

NDA 203756

Drug Name: Cabozantinib (S)- NDA/BLA Type: Priority malate

On **initial** overview of the NDA/BLA application for RTF: No

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

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? ____√____

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

NA

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.	√			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	√			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.	√			
Appropriate references for novel statistical methodology (if present) are included.			√	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	√			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	√			

File name: 5_Statistics Filing Checklist for a New NDA 203756

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

/s/

YUAN L SHEN
06/08/2012

KUN HE
06/12/2012