Cabozantinib in Progressive Medullary Thyroid Cancer

Elisei, et al

DOI: 10.1200/JCO.2012.48.4659

CLINICAL STUDY PROTOCOL

An International, Randomized, Double-Blinded, Phase 3 Efficacy Study of Cabozantinib versus Placebo in Subjects with Unresectable, Locally Advanced, or Metastatic Medullary Thyroid Cancer

1 STUDY OBJECTIVES

The primary objective of this study is:

- To evaluate PFS with cabozantinib (XL184) treatment as compared with placebo in subjects with unresectable, locally advanced, or metastatic MTC

The secondary objectives of this study are:

- To evaluate OS with cabozantinib treatment as compared with placebo
- To evaluate the objective response rate (ORR) and duration of response in subjects with measurable disease with cabozantinib treatment as compared with placebo, as per modified Response Evaluation Criteria In Solid Tumors (mRECIST)[1]
- To evaluate changes in serum levels of calcitonin and CEA as prognostic biomarkers for cabozantinib treatment benefit as compared with placebo
- To assess the potential relationship between RET germline and/or tumor DNA sequence alteration and the efficacy of cabozantinib
- To assess the pharmacodynamic effects of cabozantinib
- To evaluate the safety and tolerability of cabozantinib treatment
- To assess the PK of cabozantinib

The exploratory objective of this study is:

- To evaluate subject self-assessment parameters and symptom burden with cabozantinib treatment as compared with placebo, as per the MD Anderson Symptom Inventory (MDASI) Thyroid Module

[1] mRECIST consists of modifications provided by the Independent Radiology Review Committee that further clarifies the rules of the original RECIST 1.0 (Therasse 2000) criteria.

2 STUDY DESIGN

2.1 Study Sites

This study will be conducted at up to 140 active enrolling clinical sites in regions including, but not limited to the United States, Europe, Canada, Latin America, Asia-Pacific, and Australia.

2.2 Key Study Endpoints

The primary efficacy endpoint is duration of PFS.

The key secondary efficacy endpoints for this study are ORR in subjects with measurable disease and duration of OS.

Safety endpoints include adverse events and laboratory parameters.

2.3 Overview of Study Design

This is an international, randomized, double-blinded, multi-center, placebo-controlled Phase 3 study of unresectable, locally advanced, or metastatic MTC. Three hundred fifteen eligible subjects will be enrolled in two parallel treatment arms to receive either a single oral daily dose of cabozantinib or placebo comparator. Enrollment of subjects with only non-measurable disease will be capped at 31 (10%). Subjects will be randomized in a 2:1 ratio to receive either cabozantinib or placebo, respectively. The estimated duration of the study is 33 months for subject accrual. This study will include a Pre-Treatment Period (screening and baseline evaluations), a Treatment Period (consisting of 4-week cycles of either cabozantinib administered daily orally at a dose of 140 mg [equivalent to 175 mg malate salt] or matching placebo administered daily orally), and a Post-Treatment Period (consisting of a Post-Treatment Visit 30 days [+7] after the last dose of study treatment, with follow-up information obtained every 12 weeks [± 15 days] post last dose thereafter until final survival status is determined).

Tumor assessments will be performed every 12 weeks (± 5 days) following randomization until documented disease progression per mRECIST by the investigator. Tumor assessments will be also be evaluated by the Independent Radiology Review Committee (IRC).

2.4 Treatment Groups and Randomization

The study will include two parallel treatment arms as follows:

- Cabozantinib

- Placebo comparator

Cabozantinib or placebo will be administered daily orally at a dose of 140 mg in a double-blind fashion. Each cycle will be four weeks in duration.

Randomization will occur in a 2:1 fashion (cabozantinib: placebo) and will be stratified by the following categories:

- Age
  - ≤ 65 years
  - > 65 years

- Prior tyrosine kinase inhibitor status:
  - Known prior receipt of a tyrosine kinase inhibitor (Yes/No)

3 STUDY POPULATION

3.1 Target Population

This study will be conducted in subjects with pathologically confirmed, unresectable, locally advanced, or metastatic MTC.

3.2 Inclusion Criteria

A subject must meet the following criteria to be eligible for the study:

1. The subject has a histologically confirmed diagnosis of MTC that is unresectable, locally advanced, or metastatic, and disease that is measurable or non-measurable per mRECIST.
2. The subject is at least 18 years old.
3. The subject has an ECOG (Eastern Cooperative Oncology Group) performance status ≤ 2.
4. The subject has documented PD on computerized tomography (CT), magnetic resonance imaging (MRI), bone scan, or X-ray (determined by the investigator) per mRECIST at screening compared with a previous image done within 14 months of screening (see Section 4.4.1).[2]
5. The subject has recovered to National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v3.0 Grade ≤ 1 from clinically significant AEs due to anti-neoplastic agents, investigational drugs, or other medications that were administered prior to randomization.
6. The subject has organ and marrow function as follows: absolute neutrophil count ≥ 1500/mm^3, platelets ≥ 100,000/mm^3, hemoglobin ≥ 9 g/dL, bilirubin ≤ 1.5 times the upper limit of normal (does not apply to subjects with Gilbert's syndrome), serum

[2] Approximately 90% of subjects were enrolled under Protocol Amendment 1 which required progressive disease at study entry per the Independent Radiology Review Committee

creatinine ≤ 1.5 mg/dL, and alanine aminotransferase (ALT) and aspartate aminotransferase (AST) ≤ 2.5 times the upper limit of normal.
7. Sexually active subjects must agree to use medically accepted methods of contraception during the course of the study and for 3 months following discontinuation of study treatments (excluding women who are not of child bearing potential and men who have been sterilized).
8. The subject has no other diagnosis of malignancy (unless non-melanoma skin cancer, carcinoma in situ of the cervix, or a malignancy diagnosed ≥ 2 years previously) and currently has no evidence of malignancy (unless non-melanoma skin cancer or carcinoma in situ of the cervix).
9. Female subjects of childbearing potential must have a negative pregnancy test at screening. Females of childbearing potential are defined as sexually mature women without prior hysterectomy or who have had any evidence of menses in the past 12 months. However, women who have been amenorrheic for 12 or more months are still considered to be of childbearing potential if the amenorrhea is possibly due to prior chemotherapy, anti-estrogens, or ovarian suppression.

3.3 Exclusion Criteria

1. The subject has received prior systemic anti-tumor therapy (eg, chemotherapy, biologic modifiers, or anti-angiogenic therapy) within 4 weeks of randomization (6 weeks for nitrosoureas or mitomycin C).
2. The subject has received radiation to ≥ 25% of bone marrow.
3. The subject has received treatment with other investigational agents within 4 weeks of randomization.
4. The subject has received treatment with cabozantinib.
5. The subject has brain metastases or spinal cord compression, unless completed radiation therapy ≥ 4 weeks prior to randomization and stable without steroid and without anti-convulsant treatment for ≥ 10 days.
6. The subject has a history of clinically significant hematemesis or a recent history of hemoptysis of > 2.5 mL of red blood or other signs indicative of pulmonary hemorrhage or evidence of endobronchial lesion(s).
7. The subject has a urine protein/creatinine ratio of ≥ 1 (reported in grams of protein over grams of creatinine).
8. The subject has serious intercurrent illness, such as hypertension (two or more blood pressure readings performed at screening of > 140 mmHg systolic or > 90 mmHg diastolic) despite optimal treatment, unhealed wounds from recent surgery, or cardiac arrhythmias; or a recent history of serious disease such as either symptomatic congestive heart failure or unstable angina pectoris within the past 3 months, or myocardial infarction, stroke, or transient ischemic attack within the past 6 months.
9. The subject is pregnant or breastfeeding.
10. The subject has an active infection requiring systemic treatment.
11. The subject has a known allergy or hypersensitivity to any of the components of the cabozantinib or placebo formulations.
12. The subject is incapable of understanding and complying with the protocol or unable to provide informed consent.

4 STUDY ASSESSMENTS AND PROCEDURES

4.1 Overview

This document generally presents scheduled times for study procedures by abbreviated references to Cycle (C) and Day (D) number, as in "C1D1" where Day 1 of each cycle is the first date of study treatment for that cycle.

Cycle 1 Day 1 is defined as the first day of study treatment after randomization; cycle days are counted sequentially thereafter. Cycle 1 ends on Cycle 1 Day 28, and Cycle 2 begins the next day (Cycle 2 Day 1). This convention for determining the start and stop dates for cycles is maintained until treatment is permanently discontinued, regardless of whether treatment is interrupted or withheld for any period.

The study consists of the following periods and procedures:

Pre-Treatment Period: Subjects will undergo screening evaluations to determine study eligibility, including medical history, physical examination, hematology and serum chemistry laboratory assessments, and complete tumor assessment. Baseline subject self-assessment parameters (including assessments of pain and diarrhea) and symptom burden will be evaluated using the MDASI Thyroid Module. Screening will be conducted within 28 days of randomization. Assessments including physical examinations (including height and weight), ECOG performance status, hematology, serum chemistry, urine protein/creatinine ratio, urinalysis, and pregnancy tests will be repeated only if screening evaluations were obtained more than 7 days before C1D1. All screening and eligibility assessments must be performed prior to randomization including determination of PD (Inclusion Criterion #4) by the investigator (see Section 4.4.1.1).

Treatment Period: After completion of Pre-Treatment assessments and confirmation of study eligibility, subjects will be randomized to receive cabozantinib or placebo in a double-blind fashion and enter the Treatment Period. cabozantinib or placebo will be administered daily orally at a dose of 140 mg (one 80-mg capsule and three 20-mg capsules) in a double-blind fashion. Each cycle will be 4 weeks in duration. Study treatment will be administered by clinical site personnel at a subset of protocol-defined visits or self-administered on an outpatient basis. All protocol-specified clinic visits by subjects should occur within ± 2 days during Cycles 1 and 2 or within ± 5 days during Cycle 3 and beyond of the nominal visit day unless otherwise specified.

Tumor assessments will be performed every 12 weeks (± 5 days) following randomization until documented tumor progression by the investigator, per mRECIST. After study treatment is discontinued due to documented progression per mRECIST as assessed by the investigator or due to unacceptable toxicity, the subject enters the Post-Treatment Period. At this time, the subject is free to receive further treatment deemed necessary by their physician.

Post-Treatment Period: Subjects will return to the study site 30 days (+7) after the last dose of study treatment to complete the 30 day Post-Treatment assessment. The investigator (or designee) will obtain follow-up information (including survival status) every 12 weeks (± 15 days) post last dose thereafter until final survival status is determined. If treatment was discontinued for reasons other than PD or withdrawn consent, tumor assessments will continue every 12 weeks thereafter (± 5 days), following randomization until documented tumor progression by the investigator, per mRECIST

4.2 Randomization

After completion of Pre-Treatment procedures and confirmation of subject eligibility, including determination of PD by the investigator (Inclusion Criterion #4), the subject will be randomized into the study. A subject is defined to be enrolled in the study upon randomization. Subjects who are randomized into the study are to undergo all subsequent evaluations required by the protocol (including tumor assessments and survival follow-up), even if no study treatment is administered, until documented death or withdrawal of informed consent.

4.3 Administration of Study Treatment During the Treatment Period

Subjects will be provided with a sufficient supply of study treatment and instructions for taking the study treatment on days without scheduled clinic visits. After fasting (with exception to water) for 2 hours, subjects will take study treatment daily each morning with a full glass of water (minimum of 8 oz./ 240 mL) and continue to fast for 1 hour after each dose of study treatment. Subjects must be instructed to not make up missed doses unless the missed dose can be taken within 12 hours of the normal dosing time. Subjects must be instructed to not administer extra doses due to vomiting.

4.3.1 Extended Follow-Up Period

Every 12 weeks (± 15 days) after last dose of study treatment until final survival status is determined, the investigator (or designee) will acquire the following information from subjects who were randomized:

- Subsequent cancer treatments
- If subject died, date and cause of death

4.4 Tumor Assessment

4.4.1 Routine Tumor Assessment

Tumor response will be assessed using the mRECIST. All known lesions will be assessed. All site assessments of radiographic images will be performed by a radiologist.

4.4.1.1 Screening

Documentation of progressive disease as determined by the investigator per mRECIST is required for enrollment of a subject (inclusion criterion # 4) will be verified by the same radiologic assessment method previously used.

- A ***historical reference image*** may be performed up to 14 months before screening.
- A ***qualifying image*** will be compared to the historical reference image to establish progressive disease
- A ***screening image*** (used as the baseline for prospective evaluation) may or may not be the same as the ***qualifying image***

At screening, tumors will be assessed by bone scan (scintigraphy), liver MRI, and MRI or CT of the head, neck, chest, and abdomen, per mRECIST. Liver metastases may be assessed by contrast-enhanced triple phase CT when MRI is not possible. If lesions are seen or suspected on bone scan, an X-ray, CT, or MRI of the location of the bone scan lesion will be done.

4.4.1.2 Post-Randomization

All known lesions will be assessed. After screening, tumors will be assessed by liver MRI and MRI or CT scans of the neck, chest, and abdomen every 12 weeks (± 5 days) from randomization until PD per mRECIST. If there are bone lesions at baseline, an X-ray, CT, or MRI of the bone lesion(s) will be repeated every 12 weeks (± 5 days) until PD per mRECIST. Bone scans will only be acquired at follow up visits if clinically indicated (ie, new metastasis suspicion). If lesions are seen or suspected on the bone scan, an X-ray,

CT, or MRI of the location of the bone scan lesion will be acquired. The same radiologic assessment method will be used to assess a lesion at screening and after randomization.

Responses will be confirmed with a follow-up tumor assessment at approximately 30 days (but no earlier than 4 weeks) after the criteria for the initial response are first met. Confirmatory scans will not alter the original response evaluation schedule.

Tumor assessments will be evaluated by the investigator and by a blinded IRC. The procedures to be followed by the IRC will be defined in an IRC charter.

4.4.2 Determination of Progression

Progressive disease will be assessed using mRECIST. For purposes of subject management, including the decision to discontinue study treatment, PD will be assessed by the investigator. Detailed records of PD (including scan measurements) will be recorded in the case report forms (CRFs). Subsequently, the IRC will conduct a blinded evaluation of scans to evaluate PD and objective response for the purpose of evaluating primary and secondary study endpoints.

4.5 Patient Reported Outcomes

4.5.1 Symptom Evaluation

Each patient will be assessed (including assessments of pain and diarrhea) using the MDASI module for thyroid cancer. The MDASI Thyroid Module consists of 13 “core” cancer and treatment related symptoms (Cleeland 2000), and six additional items developed for the specific symptoms of persons with thyroid cancer (Trask 2008). The MDASI also measures how symptoms interfere with daily activities. The MDASI Thyroid Module was developed to meet concerns with PROs as outlined in the FDA Guidance for Patient Reported Outcomes (2006), including patient involvement in the generation of items, and patient debriefing of the final scale for relevance, ease of understanding, and acceptability. The MDASI Thyroid Module will be administered during a clinic visit following randomization and prior to beginning therapy.

Subject self-assessment parameters and symptom burden (including clinical symptoms such as pain, fatigue, nausea, diarrhea, and mood) will be assessed per MDASI Thyroid Module at screening and every 12 weeks (± 5 days) from randomization until PD per mRECIST , to coincide with tumor assessments.

4.6 Withdrawal Criteria from Study

Subjects may withdraw from study treatment or withdraw their consent to participate in the study at any time without prejudice. The investigator may withdraw a subject from study treatment or from the study if, in his or her clinical judgment, it is in the best interest of the subject or if the subject cannot comply with the protocol.

The investigator will also withdraw a subject from study treatment or from the study upon Exelixis' request or if Exelixis chooses to terminate the study.

In addition, any of the following conditions require withdrawal of the subject from study treatment:

- Adverse event or intercurrent illness that in the opinion of the investigator warrants the subject's withdrawal from treatment
- Progressive disease per mRECIST as determined by the investigator (the determination of PD should only be made based on unequivocal evidence of progressive disease)
- Necessity for treatment with other investigational drug or other anti-cancer medications prohibited by protocol
- Participation in another clinical study using an investigational agent
- Significant noncompliance with the protocol schedule as deemed by the investigator or Sponsor
- Request by regulatory agencies for termination of treatment of an individual subject or all subjects under this protocol
- Sexually active subjects (excluding women who are not of child bearing potential and men who have been sterilized) who refuse to use medically accepted methods of contraception during the course of the study and for 3 months following discontinuation of study treatment
- Subject unable to tolerate a minimum dose of 60 mg (see Table 4-1).

4.7 Dose Modifications and Delays

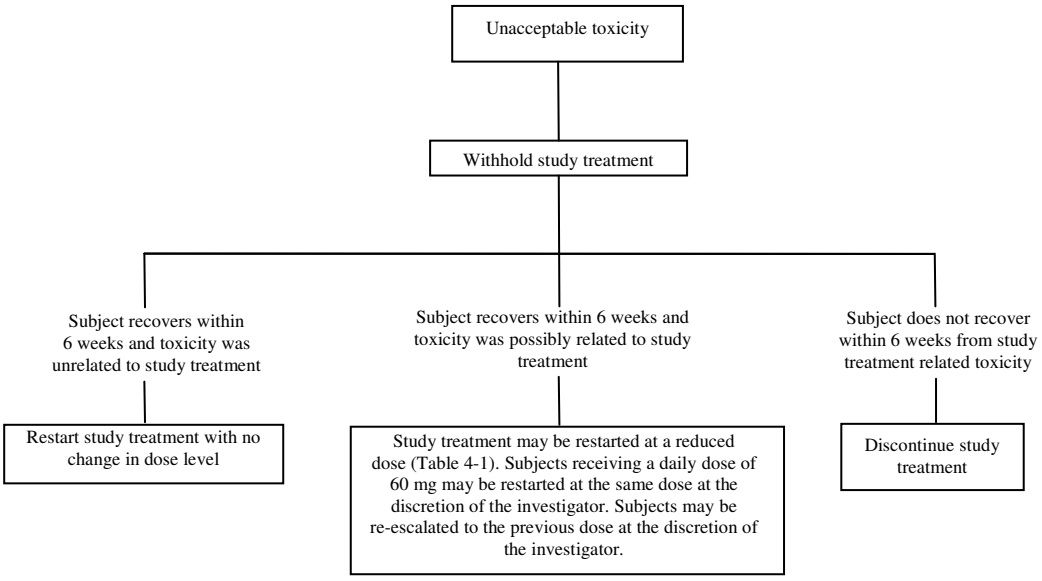
4.7.1 Dose Reduction or Treatment Delay for Toxicity

Subjects will have treatment withheld if they develop unacceptable toxicity defined as:

- Intolerable Grade 2 toxicity that cannot be adequately managed
- Grade 3 or greater non-hematological toxicity (including nausea, vomiting, or diarrhea despite optimal management)
- Urine protein/creatinine ratio > 2
- Grade 4 hematological toxicity

The schema for treatment delays and reductions in response to unacceptable toxicity experienced by individual subjects is shown in Figure 4-1.

Figure 4-1: Dose Modification Schema



If a subject experiences unacceptable study treatment related toxicity (per the criteria above), study treatment may be withheld at the investigator's discretion if clinically necessary. If the subject does not recover from his or her toxicities to Grade ≤ 1 or to the baseline value (or lower) after 6 weeks, the subject will receive no further study treatment.

If the subject recovers from his or her toxicities to Grade ≤ 1 or to the baseline value (or lower) within 6 weeks (per the criteria above) and the toxicity was unrelated to study treatment, then study treatment may be restarted with no change in dose.

If the subject recovers from his or her toxicities to Grade ≤ 1 or to the baseline value (or lower) within 6 weeks (per the criteria above) and the toxicity was deemed possibly related to study treatment, then study treatment may be restarted at a reduced dose (see Table 4-1 for the schedule of dose reductions). Subjects receiving a daily dose of 60 mg may be restarted at the same dose at the discretion of the investigator. Subjects may be re-escalated to the previous dose at the discretion of the investigator no sooner than 2 weeks beyond resolution (Grade ≤ 1 or to the baseline value [or lower]) of symptoms that

led to the dose reduction. Subjects unable to tolerate a dose of 60 mg (see Table 4-1) will discontinue study treatment.

If study treatment is withheld, the subject should be instructed not to "shift" or make up the withheld doses unless the missed dose can be taken within 12 hours of the normal dosing time.

- For example, if doses are withheld on Days 11-15 of a 28-day daily-dosing cycle, dosing would resume (if indicated) on Day 16 and the cycle would end as scheduled on Day 28.
- If doses are withheld on Days 1-4 of a 28-day cycle, dosing would resume (if indicated) on Day 5 and the cycle would end as scheduled on Day 28. (Except for Cycle 1, wherein the first day of Cycle 1 is defined as the date the first dose of study treatment is administered).

Table 4-1: Dose Reductions

First Dose Level Reduction	Second Dose Level Reduction
Cabozantinib 100 mg or matching placebo oral qd	Cabozantinib 60 mg or matching placebo oral qd

5 STATISTICAL CONSIDERATIONS

Details of the planned analyses will be provided in a separate Statistical Analysis Plan (SAP) that will be finalized before the last subject is randomized. The statistical principles applied in the design and planned analyses of this study are consistent with ICH E9.

5.1 Analysis Populations

The following populations will be employed for statistical analyses.

5.1.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population will consist of all subjects who are randomized, regardless of whether any study treatment or the correct study treatment is received.

5.1.2 Safety Population

The Safety population will consist of all subjects who receive any amount of treatment. Subjects who receive both treatments in error will be summarized in the cabozantinib group.

5.1.3 Per Protocol Population

The Per Protocol (PP) population will consist of all subjects in the Safety population who

- Have a baseline and at least one post-baseline blinded tumor assessment
- Experience no major protocol violations (to be defined in the SAP)

5.2 Primary Efficacy Endpoint

The primary efficacy endpoint is duration of PFS.

5.2.1 Definition

Duration of PFS is defined as the time from randomization to the earlier of the following events: documented PD per mRECIST, or death due to any cause.

Censoring rules for the primary analysis are described in Section 5.2.2.

5.2.2 Primary Analysis

The primary analysis of PFS will be performed using the ITT population. It is designed to include only progression events as determined by the IRC per mRECIST. Clinical deterioration or radiographic progression determined by the investigator will not be considered progression events.

- Subjects who do not have any post-baseline tumor assessments will be right censored on the date of randomization.
- Subjects who receive subsequent anti-cancer therapy before experiencing an event will be right censored at the date of the last tumor assessment prior to the date of initiation of subsequent therapy.
- 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 tumor assessment

The primary analysis of PFS is event-based and will be conducted after at least 315 subjects have been randomized (or the study has otherwise been closed to accrual) and at least 138 events (progression per mRECIST as assessed by the IRC or death) have been observed in the study.

Hypothesis testing between the two treatment arms will be performed using the stratified log-rank test with a 2-sided 0.05 level of significance. The stratification factors will be:

- Age
    - $\leq$ 65 years
    - > 65 years

- Prior tyrosine kinase inhibitor status
    - Known prior receipt of a tyrosine kinase inhibitor (Yes/No)

The median duration of PFS and the associated 95% confidence interval for each treatment arm will be estimated using the Kaplan-Meier method. The hazard ratio (HR) will be estimated using a Cox regression model and will include the same stratification factors described above.

In the primary analysis of PFS, if the p-value for the stratified log-rank test is statistically significant and the HR ($\lambda_{cabozantinib}/ \lambda_{placebo}$) is < 1, the null hypothesis of no difference in PFS will be rejected and it will be inferred that PFS is superior in the group receiving cabozantinib compared with the group receiving placebo.

5.2.3 Supportive Analyses

Supportive (sensitivity) analyses of PFS will be defined in the SAP using alternative event definitions and censoring schemes to account for partial or completely missing assessments, address bias due to tumor assessment timing, and to address potential discrepancies between the documentation of progression per the investigator and per the IRC. These analyses will be performed using the same statistical methods described for the primary efficacy analysis. Analyses using the Per Protocol population will also be performed\.

Exploratory analyses of the effect of baseline characteristics, stratification factors, and other variables will be conducted using Cox regression models and subgroup analyses performed employing Kaplan-Meier methods.

5.3 Secondary Efficacy Endpoints

5.3.1 Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints for this study are duration of OS and ORR. Formal hypothesis testing is planned for key secondary efficacy endpoints.

5.3.1.1 Overall Survival

Duration of OS is defined as the time from randomization to death due to any cause. For subjects who are alive at the time of data cutoff or are permanently lost to follow-up, duration of OS will be right censored at the date the subject was last known to be alive.

The primary analysis of OS is event-based and will be conducted after at least 217 deaths have been observed in the study.

Statistical methods for the primary, supportive, and exploratory analyses of OS will be applied as described for the primary endpoint of PFS.

An interim analysis of OS, to be performed at the time of the primary analysis of PFS, is described in Section 5.6.

5.3.1.2 Objective Response Rate

The ORR is defined as the proportion of subjects who have measurable disease at baseline for whom the best overall response at the time of data cutoff is complete response (CR) or PR as assessed by the by the IRC per mRECIST, which is confirmed by a subsequent visit $\geq$ 28 days later.

Hypothesis testing will be performed using the chi-squared test at the 2-sided 0.01 α level.

Point estimates of ORR, the difference in response rates between the two treatment arms, and associated confidence intervals will be provided. Confidence intervals will be calculated using exact methods.

The primary analysis of ORR will be performed at the time of the primary analysis of the primary endpoint, based upon response as determined the by IRC and using the ITT population. Subjects who do not have any post-baseline blinded tumor assessments will be counted as non-responders.

Supportive statistical analyses of ORR will be performed in the PP population and of response based upon investigator assessment (in both the ITT and PP populations).

If sufficient responses are observed, additional supportive analyses will be conducted using appropriate methods to adjust for stratification factors.

5.3.1.3 Control of Type I Error

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) will be addressed by employing a fixed-sequence testing procedure, dividing the alpha between the secondary endpoints, and implementing an alpha-spending function.

When the required number of events has been reached for the primary endpoint, the hypothesis for PFS will be tested at the 2-sided 0.05 (α) significance level. If this hypothesis is rejected at this α level, the two key secondary endpoints will then be tested

in parallel. Objective response rate will be tested at the 2-sided 0.01 α level, and an interim analysis of OS will be performed at the 2-sided 0.00006 level per a Lan-DeMets O'Brien-Fleming alpha-spending function.

If the result of the interim analysis for OS is not significant, the final analysis of OS will be performed when the required number of deaths has occurred, at the 2-sided 0.04 significance level (total 2-sided alpha spent for OS of 0.04) per the alpha-spending function.

All other statistical evaluations of efficacy will be considered exploratory.

5.3.2 Supportive Secondary Efficacy Endpoints

Duration of response is a supportive secondary efficacy endpoint for this study. No formal hypothesis tests are planned for duration of response.

5.3.2.1 Duration of Response

Duration of response is defined as the time from first documentation of objective response that is subsequently confirmed at a visit that is $\geq$ 28 days later to PD by mRECIST or death due to any cause. Responders who have not been documented to have progressed or died at time of data cutoff will be right censored at the last available blinded tumor assessment.

Median duration of response for each treatment arm and its associated 95% confidence interval will be estimated using the Kaplan-Meier method.

5.4 Patient Reported Outcomes

Patients will be assessed (including assessments of pain and diarrhea) using the MDASI module for thyroid cancer. The MDASI Thyroid Module consists of 13 "core" cancer and treatment related symptoms (Cleeland 2000) and six additional items developed for the specific symptoms of persons with thyroid cancer (Trask 2008). The MDASI also measures how symptoms interfere with daily activities.

For this study, the primary severity score of the MDASI Thyroid Module is the mean symptom score defined as the average of a subject's responses to the following items: diarrhea, fatigue, sleep disturbance, distress, and difficulty remembering.

Additional scores include:

- Overall mean symptom severity score is defined as the average of a subject's MDASI Thyroid Module responses (13 "core" and 6 specific thyroid items).
- Mean symptom interference subscale score is defined as the average of a subject's six interference items.

Details of the planned analyses for these outcomes will be provided in the SAP.

5.5 Safety Analyses

All safety analyses will be performed using the Safety population. No formal comparisons between the two treatment arms are planned.

5.5.1 Adverse Events

Adverse event terms recorded on the CRFs will be mapped to preferred terms using the MedDRA dictionary. The investigator will classify the severity of AEs using the CTCAE v3.0 and will judge each event to be "not related" or "possibly related" to study treatment.

A treatment emergent adverse event (TEAE) is defined as any event with an onset date on or after date of first dose of study treatment, or any event present before treatment that worsens after treatment. Only TEAEs with an onset date prior to date of last dose + 30 days will be tabulated in summary tables.

The frequency and percentage of subjects with TEAEs will be tabulated for overall incidence by system organ class and preferred term by treatment arm. Related TEAEs, serious TEAEs, related serious TEAEs, and TEAEs resulting in study treatment discontinuation will be similarly summarized. TEAEs and related TEAEs will also be summarized for worst reported severity within each subject.

At each level of summarization, a subject will be counted only once for each AE preferred term he/she experiences within that level (ie, multiple episodes of events with the same preferred terms will be counted only once).

All reported subject deaths will be summarized by treatment group, cause of death, and relationship to study treatment.

5.5.2 Laboratory Test Results

Continuous laboratory test results will be summarized by treatment group using descriptive statistics for actual values and for changes from baseline.

Shift from baseline in CTCAE grade (where applicable) and by high/low flags (where CTCAE grades are not defined) will be presented by treatment group and visit.

5.5.3 Other Safety Endpoints

Vital signs will be summarized by treatment group using descriptive statistics for actual values and for changes from baseline.

The ECOG Performance Status will be summarized as shift from baseline by treatment group and visit.

QTc interval will be summarized by treatment group using descriptive statistics for actual values and for changes from baseline.

Shift in QTc CTCAE grade from baseline by treatment group and visit will be also presented.

Concomitant medications will be standardized using the World Health Organization drug dictionary and summarized by class and preferred term.

5.6 Interim Analyses

A single interim analysis for the key secondary efficacy endpoint of OS is planned. The interim analysis will be conducted at the time of the primary analysis of PFS. It is anticipated that this will be at approximately the 31% information fraction for OS (eg, after approximately 67 deaths have been observed). Type I error for the interim analysis will be controlled by implementing a Lan-DeMets O'Brien-Fleming alpha-spending function as described in Section 5.3.1.3.

5.7 Power and Sample Size

The study is designed to provide adequate power for both PFS and OS.

For the primary endpoint of PFS, assuming exponential PFS, proportional hazards, and a 2:1 treatment allocation ratio (cabozantinib:placebo), 138 events are required to provide 90% power to detect a HR of 0.571 using the log-rank test and a 2-sided significance level of 5%.

This corresponds to a 43% reduction in the risk of progression or death, or a 75% improvement in median PFS from 8 months to 14 months.

Under this design, the minimum observed effect that would result in statistical significance for PFS is a 40.3% improvement (HR = 0.713) in PFS from 8 to 11.2 months. If the true treatment effect is a 75% improvement in progression-free survival, there is a 90% chance (power) of observing a 40.3% or greater improvement, and a 50% chance of observing a 75% or greater improvement.

For the key secondary efficacy endpoint of OS, assuming a single interim analysis at the 31% information fraction (at the time of the primary analysis of PFS) and a subsequent primary analysis, 217 deaths are required to provide 80% power to detect a HR of 0.667 using the log-rank test and a 2-sided significance level of 4%.

This corresponds to a 33.3% reduction in the risk of death, or a 50% improvement in median survival from 22 to 33 months.

Under this design, the minimum observed effect that would result in statistical significance for the primary analysis of OS is a 33.3% improvement (HR=0.750) in OS from 22 to 29.3 months. If the true treatment effect is a 50% improvement in survival, there is an 80% chance (power) of observing a 33.3% or greater improvement, and a 50% chance of observing a 50% or greater improvement.

A total of 315 eligible subjects are expected to be randomized and followed (210 to cabozantinib and 105 to placebo) 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).

Power and sample size estimates were estimated using EAST v5 by Cytel Software.

5.8 Pharmacokinetic Analysis

The concentration of study treatment in plasma will be analyzed using a validated bioanalytical method. Biotransformation products of cabozantinib may also be evaluated in plasma samples collected during the study. Descriptive statistics (including number, mean and/or median, standard deviation, and coefficient of variation) will be used to describe the concentration-time data. Where appropriate, these data may be combined with data from other studies as part of a meta-analysis. The influence of exposure on clinical safety parameters (eg, QTc, selected adverse events) will also be explored. The

results of the PK analysis will be evaluated in conjunction with available safety and pharmacodynamic data.

Bioanalytical personnel for PK, pharmacodynamic, and tissue sample analysis not involved in the conduct of the study may be unblinded as a result of the analysis. Data from these analyses will be communicated to the pharmacokineticists without disclosing subject identification.

5.9 Pharmacodynamic Analysis

5.9.1 Molecular Analysis of Blood and/or Tumor Tissue Samples

Tumor samples will be collected for pharmacodynamic analysis where not prohibited by local regulation.

Blood and tumor tissue samples will be analyzed for tumor-relevant mutations, in particular the mutational status of the RET RTK gene. Blood samples will reveal any potential RET mutations in subjects with hereditary MTC while tumor samples will allow assessment of RET mutations in subjects with either hereditary or sporadic MTC. The mutational status will not be used to stratify enrollment or to conduct the primary analyses, but is part of the exploratory analysis as outlined in the SAP.

Bioanalytical personnel for blood and/or tissue sample analysis not involved in the conduct of the study may be unblinded during the course of the analysis. Data from these analyses will be communicated to the translation medicine representatives without disclosing subject identification.

5.9.2 Pharmacodynamic Plasma Marker Analysis

Plasma samples will be collected for potential analysis of cabozantinib mechanism-of-action-based biomarkers as well as exploratory studies to determine clinical response biomarkers. Descriptive statistics will be used to describe the concentration-time data and to analyze relative changes from baseline (such as paired t-test). Where appropriate, the data may be combined with data from other studies as part of a meta-analysis. The results of the pharmacodynamic analysis may be evaluated in conjunction with available PK and safety data.

Bioanalytical personnel for pharmacodynamic sample analysis not involved in the conduct of the study may be unblinded during the course of the analysis. Data from these

analyses will be communicated to the translation medicine representatives without disclosing subject identification.

5.10 Pharmacogenomic Analysis

Blood samples will be collected for assessment of mutational status of RET and to correlate drug response or toxicity with human DNA sequence variation. The DNA sample will not be utilized for diagnostic genetic testing (eg, Tay Sachs, cystic fibrosis, HNPCC, Huntington's Disease) and the Sponsor will be blinded as to the subjects identity. The germline RET status of each subject will be performed with research grade sequencing and is intended for study analysis of clinical outcome, not for genetic counseling purposes, thus this information will not be released to the investigators and/or the subject or his/her relatives.

Bioanalytical personnel for pharmacodynamic sample analysis not involved in the conduct of the study may be unblinded during the course of the analysis. Data from these analyses will be communicated to the translation medicine representatives without disclosing subject identification.

6 REFERENCES

Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, Engstrom MC. Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. Cancer 2000;89(7):1634-1646.

Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures. FDA 2006. Available from http://www.fda.gov/CDER/GUIDANCE/5460dft.pdf. Accessed 08 April 2008.

Therasse P, Arbuck SG, Eisenhauer, EA, et al: New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 92:205–216, 2000.

Trask P, Gning I, Mendoza T, Harle M, Gutierrez K, Kitaka S, Sherman S, & Cleeland C. Initial development and validation of the thyroid cancer module of the M.D. Anderson Symptom Inventory. Psycho-Oncology 2008;17:S99.