STATISTICAL ANALYSIS PLAN

Calithera Biosciences, Inc.

CX-839-008

Protocol Title:	A Randomized, Double-Blind, Placebo-Controlled Phase 2 Clinical Trial Comparing CB-839 in Combination with Cabozantinib (CB-Cabo) vs. Placebo with Cabozantinib (Pbo-Cabo) in Patients with Advanced or Metastatic Renal Cell Carcinoma (RCC)
Sponsor:	Calithera Biosciences, Inc. 343 Oyster Point Blvd, Suite 200 South San Francisco, CA 94080
Prepared By:	[redacted]
Document Version and Date:	Version 1.0; 10 June, 2020

[redacted]

I confirm that I have reviewed this document and agree with the content.

[redacted]	
[redacted]	Signature:
[redacted]	Signature:
[redacted]	Signature: Date:
[redacted]	Signature: Date:

I confirm that I have reviewed this document and agree with the content.

APPROVALS	
Authored by: [redacted] Director of Biometrics Calithera Biosciences, Inc.	Signature: Date:
Approved by: [redacted] FMD K&L	Signature: Date:
Approved by:	Signature: Date:
Approved by:	Signature: Date:

1 TABLE OF CONTENTS

TABLE OF TABLES

2 LIST OF ABBREVIATIONS

Table 1 List of Abbreviations

Abbreviation	Definition
AE	adverse event
ATC	Anatomical Therapeutic Chemical
BID	twice daily
BMI	body mass index
CB-Cabo	CB-839 in combination with Cabozantinib
ccRCC	clear cell renal cell carcinoma
CR	complete response
CRF	case report form
CSR	clinical study report
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DCR	disease control rate
DOR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EQ 5D-5L	EuroQol 5D
FA	final analysis
FACT	Functional Assessment of Cancer Therapy
FKSI-19	Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index
IA	interim analysis
IDMC	Independent Data Monitoring Committee
INV	Investigator
IRC	Independent Radiology Committee
IWRS	interactive web response system
IxRS	interactive voice/web response system
KPS	Karnofsky Performance Status
MedDRA	Medical Dictionary for Regulatory Activities
mRCC	metastatic renal cell carcinoma
MRI	magnetic resonance imaging
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
ORR	overall response rate
OS	overall survival
Pbo-Cabo	placebo in combination with Cabozantinib
PD	progressive disease
PDn	pharmacodynamics
PD-1	programmed cell death protein 1
PD-L1	programmed death ligand 1

Abbreviation	Definition
PFS	progression-free survival
PK	pharmacokinetics
PR	partial response
QD	once daily
RCC	renal cell carcinoma
RECIST	response evaluation criteria in solid tumors
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SI	Système International
TEAE	treatment-emergent adverse event
TKI	tyrosine kinase inhibitor

3 INTRODUCTION

The statistical analysis plan (SAP) provides details of the planned analyses and statistical methods for the study CX-839-008 (A Randomized, Double-Blind, Placebo-Controlled Phase 2 Clinical Trial Comparing CB-839 (Telaglenastat) in Combination with Cabozantinib [CB-Cabo] vs. Placebo with Cabozantinib [Pbo-Cabo] in Patients with Advanced or Metastatic Renal Cell Carcinoma [RCC]). The background and rationale for the study can be found in the study protocol. Statistical analyses including endpoints in the SAP supersedes the protocol.

4 STUDY DESIGN

Protocol CX-839-008 is a Phase 2 randomized, double-blind, placebo-controlled, multicenter study comparing two treatment regimens for patients with clear cell RCC (ccRCC). Figure 1: Study Design SchemaEligible patients were randomized in a 1:1 ratio to the following study treatment arms:

- CB-Cabo - telaglenastat (800 mg twice daily [BID]) + cabozantinib (60 mg once daily [QD])
- Pbo-Cabo - Placebo + cabozantinib (60 mg QD)

Randomization will be stratified by the following variables:

- Prior treatment with PD-1/PD-L1 inhibitor therapy (yes vs. no)
- The International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) Prognostic Risk Group (favorable vs. intermediate vs. poor)

Treatment administrations are per the instructions and schedule of events outlined in the protocol. Crossover between treatment arms will not be allowed.

Radiographic evaluation of tumor burden (e.g., diagnostic computed tomography [CT] scan with intravenous contrast or MRI) will occur during Screening (within 28 days prior to randomization), every 8 weeks after C1D1 for the first 12 cycles, every 12 weeks beginning with cycle 13, and at the End of Treatment Visit. Radiographic assessments may occur more frequently as clinically indicated.

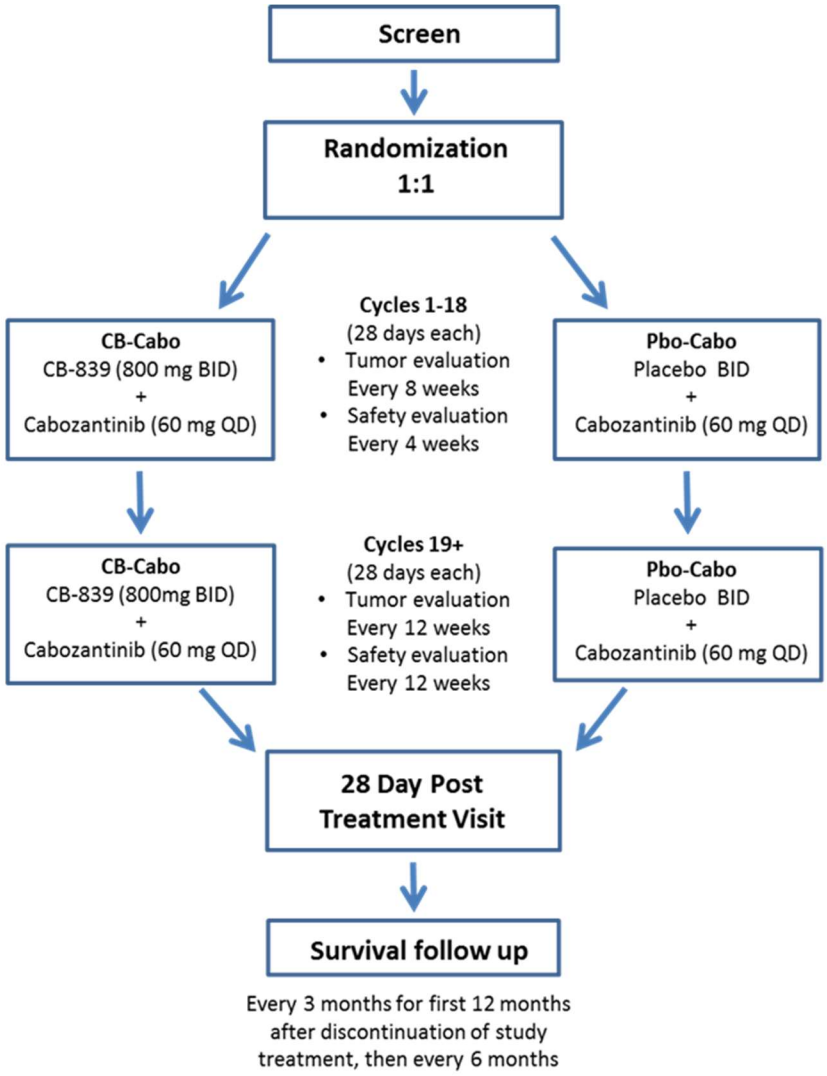
Patients will receive study treatment determined by randomization in 28-day cycles until disease progression per Response Evaluation Criteria in Solid Tumors (RECIST) v.1.1 or unacceptable toxicity, whichever occurs first. Patients will be followed for safety for at least 28 days after the last dose of all study treatments (telaglenastat/placebo and cabozantinib), or until initiation of a new anticancer therapy, if earlier. Patients who discontinue study treatment for reasons other than disease progression by RECIST v.1.1 or death will also continue to be followed by radiographic tumor imaging until documentation of radiographic progressive disease per RECIST v.1.1, death, initiation of a new anticancer therapy, or withdrawal of consent. Long-term follow up for survival

will continue until death or withdrawal of consent for survival follow-up. Patients who continue treatment beyond progression will continue to receive all scheduled protocol assessments, including continued radiographic tumor assessments, until study treatment is discontinued.

An independent Data Monitoring Committee (IDMC) monitors safety data. See Section 5.2 for more information.

Figure 1 illustrates the study design.

Figure 1: Study Design Schema



4.1 Study Endpoints

4.1.1 Efficacy Endpoints

4.1.1.1 Primary Efficacy Endpoint

Independent Radiology Committee (IRC)-assessed progression-free survival (PFS) per RECIST v1.1, defined as the time from randomization to documented disease progression or death due to any cause, whichever occurs first.

4.1.1.2 Secondary Efficacy Endpoints

Overall Survival (OS) is defined as the time from randomization to death due to any cause.

Investigator (INV)-assessed PFS per RECIST v1.1, defined as the time from randomization to documented disease progression or death due to any cause, whichever occurs first.

4.1.1.3 Additional Efficacy Endpoints

The additional efficacy endpoints are:

- Overall response rate (ORR), defined as the percentage of patients with complete response (CR) or partial response (PR) according to the RECIST 1.1 criteria per IRC
- Duration of response (DOR), defined as the time between the first documentation of a PR or a CR to the first documentation of progressive disease (PD) or death, whichever occurs first as determined by RECIST 1.1 criteria per IRC
- Disease control rate (DCR), defined as the summed percentage of patients with CR, PR and stable disease (SD) following treatment initiation according to RECIST 1.1 criteria per IRC.
- PFS rates according to RECIST v1.1 criteria per IRC at 6-month, 9-month and at 1-year landmark timepoints
- OS rate at 1-year, 2-year, and 3-year landmark timepoints
- PFS according to RECIST v1.1 criteria per IRC and OS in subgroups based on demographics and baseline characteristics
- Patient reported outcomes using the NCCN-Functional Assessment of Cancer Therapy (FACT) Kidney Symptom Index (FKSI-19) questionnaire and the EuroQoL 5D utility score (EQ-5D-5L).

- Genetic variants and other biomarkers related to angiogenesis and metabolic pathways

4.1.2 Pharmacokinetic (PK) Endpoints

- Potential relationship between drug and metabolite exposure and the efficacy and safety as well as evaluate the possible effect of telaglenastat on cabozantinib exposure

Additional PK may be summarized in a separate report and may be outside the scope of the CSR.

4.1.3 Safety Endpoints

The safety endpoints are the following:

- Type, incidence, severity, seriousness, and study drug-relatedness of adverse events graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.03 (NCI CTCAE v4.03)

- Changes in vital signs, ECGs, physical findings, and clinical laboratory test results and following study treatment administration

4.2 Determination of Sample Size

The study anticipated approximately 416 patients, with 1:1 for CB-Cabo vs. Pbo-Cabo, and is anticipated to provide sufficient efficacy data to demonstrate a clinically and statistically significant benefit for telaglenastat for the primary endpoint of PFS per IRC.

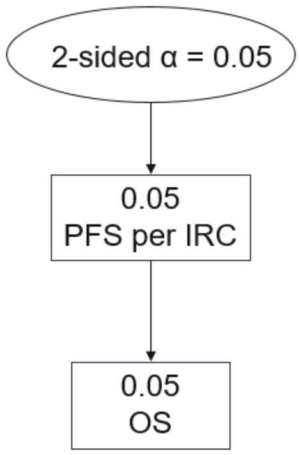
For the purposes of study design, it is assumed that the control arm's true median PFS will be approximately 8 months based upon recently published randomized Phase 3 study results with cabozantinib monotherapy (Choueiri 2015), and that the assumed true hazard ratio (HR) favoring CB-Cabo over Pbo-Cabo will be 0.69, corresponding to an increase of 3.6 months from 8 months to 11.6 months.

The proposed study enrolled patients with an accrual period of 16 months and additional follow-up of at least 8 months. With the control arm's median PFS being 8 months and the true PFS HR of experimental patients to control patients is 0.69, the study will require approximately 262 PFS events in order to reject the null hypothesis of equal PFS between the two treatment arms with a power of 0.85. This estimation assumes a 2-sided type I error rate of 0.05, uniform accrual, exponential PFS distributions and the log-rank test for PFS comparison. The estimate of the number of events required to demonstrate efficacy with regard to PFS is shown in Table 2.

To control the overall type I error rate for treatment comparisons at 2-sided test at 0.05, when PFS per IRC is statistically positive, all of the α will be passed to the secondary endpoint of OS, which will be tested at a 2-sided α level of 0.05 (see Figure 2).

OS will be tested using the group sequential method at the interim (when PFS is assessed as described above) and final analyses based on the 2-sided α level of 0.05. Statistical significance at the interim analysis of OS will be tested as described in Table 3, which also shows estimated number of OS events. The stopping boundaries for the interim and final OS analyses will be calculated using the Lan-DeMets approximation to the O'Brien-Fleming boundary. Table 3

Figure 2: Progression-Free Survival and Overall Survival Analysis Testing Strategy, Alpha Allocation



IRC, independent radiology committee; OS, overall survival; PFS, progression-free survival.

4.3 Analysis Timing

No interim analyses are planned for PFS in this study; one interim analysis is planned for OS. Because the cabozantinib median PFS is expected to be 8 months, the sponsor anticipates that having a minimum follow-up period of 8 months will provide sufficient maturity of the PFS data. The PFS primary analysis will be conducted when approximately 262 IRC-determined PFS events have occurred in the ITT population (analysis populations are defined in Section 6.1) and at least 8 months elapse after the last patient is enrolled, whichever occurs later. The expected timing for the IRC-determined PFS analysis is shown in Table 2.
At the time of the PFS primary analysis, the interim analysis of OS will be performed. The expected timing of the final analysis of OS will be determined by the number of OS events that are observed in the ITT population as shown in Table 3. The estimated number of events for the OS at the interim analysis is also shown in Table 3.

For OS analyses (IA and FA), actual boundaries will be adjusted according to the true number of OS events observed at the data cuts.

Table 2: The Expected Analysis Timing for Progression -Free Survival in the Intent-to-Treat Population

Type of Analysis	Months from FPI	Criteria	
PFS Primary	~25	Number of Events/Event Ratio	262/63%
		MDD in HR	≤0.784
		p-value stopping boundary	≤0.05
		Power	85%

~, approximately; FPI, first patient initiated screening; PFS, progression-free survival; HR, hazard ratio; MDD, minimum detectable differences. Target HR assumption is 0.69. p-values are 2-sided

Table 3: The Expected Analysis Timing for Overall Survival in the Intent-to-Treat Population

Type of Analysis	Months from FPI	Percent Information	Criteria	
OS IA	~25	45%	Estimated Events/Event Ratio	135/32%
			MDD in HR	≤0.582
			p-value stopping boundary	0.002
			Power	10.4%
OS FA	~53	100%	Events/Event Ratio	300/72%
			MDD in HR	≤0.797
			p-value stopping boundary	0.049
			Total power	80%

~, approximately; FPI, first patient initiated screening; IA, Interim analysis; FA, final analysis; HR, hazard ratio; MDD, minimum detectable differences; target HR assumption is0.723. OS interim boundary is determined by O'Brien-Fleming Method; p-values are 2-sided

5 STUDY CONDUCT

5.1 Independent Radiology Committee (IRC)

An IRC has been established to evaluate tumor scans and relevant history data of trial patients in a central, blinded, and independent fashion. The IRC is comprised of board-certified radiologists who use RECIST v1.1 to assess radiographic disease response and progression following randomization. The IRC will evaluate study data at the designated primary analysis. Additional details regarding IRC member qualification, training, methods, procedures, and other issues relevant to committee operations is described in the IRC Charter.

5.2 Data Monitoring

An Independent Data Monitoring Committee (IDMC) has been established to monitor the unblinded safety and progress of the study on a regular basis. The committee operates independently from the Sponsor and the clinical investigators. To minimize the potential introduction of bias, these individuals do not have any direct contact with the study site personnel or patients. The IDMC convenes periodically (approximately every 3 -6 months depending on the progress of the study). Details of the composition, role, responsibilities and operational considerations are provided in the IDMC Charter.

In addition to real time medical review of emergent serious adverse events (SAEs), a cross functional sponsor safety review team performs regular periodic aggregate data reviews across all telaglenastat studies.

6 STATISTICAL METHODS

6.1 Analysis Populations

The analysis populations are defined as follows:

- The ITT population is defined as all randomized patients, whether or not the patients receive the assigned treatment per the interactive web response system (IWRS). Patients in the ITT set will be analyzed according to the treatment group to which they are randomized regardless of post randomization protocol deviations. The ITT patient set will be the basis for the efficacy analysis for this study unless otherwise specified.

- Safety analysis population includes all patients who receive any study-specific treatment (telaglenastat/placebo or cabozantinib) after randomization and will be analyzed according to the actual treatment received. Patients having taken any telaglenastat tablet will be analyzed in the telaglenastat with cabozantinib combination arm. Similarly, patients randomized to the telaglenastat with cabozantinib combination arm without receiving any telaglenastat will be analyzed in the placebo plus cabozantinib arm.

- PK analyses will be based on PK observations from all patients who receive telaglenastat or cabozantinib treatment and who provide at least one evaluable PK sample.

- PRO-evaluable population is defined as ITT with a non-missing baseline assessment and at least one non-missing post-baseline assessment until treatment discontinuation

6.2 General Statistical Methods

Baseline values are defined as the last available data obtained prior to the patient receiving the first dose of any study treatments on Cycle 1 Day 1 visits unless otherwise noted. Descriptive statistics (mean, median, SD, range) will be presented for continuous variables, and frequencies and percentages will be presented for categorical variables.

6.3 Analysis of Study Conduct

Enrollment, major protocol deviations (including major deviations of inclusion and/or exclusion criteria), and reason for discontinuation from the study will be summarized by treatment arm for the ITT population. Study treatment administration and reasons for discontinuation from either component of the study treatment will be summarized by treatment arm for the safety population.

6.4 Analysis of Treatment Group Comparability

Demographic and baseline characteristics, including age, sex, race, ethnicity, baseline medical history, prior medications, disease characteristics including previous treatment, histologic subtype, current disease status, and stratification factors will be summarized by treatment arm and overall subjects combined for the ITT and safety populations.

6.5 Efficacy Analysis

6.5.1 Primary Analysis of Progression-Free Survival

Progression-free survival is defined as the time from randomization to the occurrence of disease progression as assessed by the IRC using RECIST v1.1 or death from any cause, whichever occurs first. Patients who have not experienced disease progression or death at the time of analysis of PFS will be censored at the date of the last evaluable radiographic disease assessment. The duration of primary analysis of PFS will be censored as outlined in Table 4. The sensitivity analysis will be performed if > 5% of patients have a PFS event immediately after missing at least two scans consecutively before the data cutoff, by considering such as event at the first date of documented progression or death.

Treatment comparisons will be based on the stratified log-rank test in the ITT population. The stratification factors will be those used for randomization and will be obtained from the IWRS (prior PD-1/PD-L1 inhibitor therapy [yes vs. no] and IMDC prognostic risk group [favorable vs. intermediate vs. poor]). A 2-sided $p \leq 0.05$ in favor of the CB-Cabo experimental treatment will be regarded as a positive result for the study. The null and alternative hypotheses can be phrased in terms of the survival functions $S_{PFS\_CB\text{-}Cabo}(t)$ and $S_{PFS\_Pbo\text{-}Cabo}(t)$ in arm CB-Cabo and arm Pbo-Cabo, respectively:

$$H_0: S_{PFS\_CB\text{-}Cabo\ (t)} = S_{PFS\_Pbo\text{-}Cabo\ (t)} \text{ versus } H_1: S_{PFS\_CB\text{-}Cabo\ (t)} \neq S_{PFS\_Pbo\text{-}Cabo\ (t)}$$

Kaplan-Meier methodology will be used to estimate median PFS for each treatment arm and construct survival curves for the visual description of the difference between the treatment arms. The Brookmeyer-Crowley methodology will be used to construct the 95% CI for the median PFS for each treatment arm (1).

The hazard ratio (HR), $\lambda_{PFS\_CB\text{-}Cabo}/\lambda_{PFS\_Pbo\text{-}Cabo}$, where $\lambda_{PFS\_CB\text{-}Cabo}$ and $\lambda_{PFS\_Pbo\text{-}Cabo,}$ represents the hazard of PFS in arm CB-Cabo and arm Pbo-Cabo, respectively. HR will be estimated with a stratified Cox proportional hazards regression model and the same stratification variables used for the stratified log-rank test. In addition, as a sensitivity

analysis for the primary endpoint, the above analyses will be repeated with only two levels, i.e., favorable vs. intermediate/poor, for the IMDC Prognostic Risk grouping.

To prevent a stratification cell having too few events, if a stratification cell has less than 10 PFS events (two arms combined), the stratification factor that contributes the least number of events for that cell will be removed. Stratification factors for the analyses will be removed until all stratification cells have at least 10 PFS events in two arms combined.

As a sensitivity analysis, the above analyses will be performed using the number of prior TKI cancer therapy (0 vs. ≥ 1) and IMDC prognostic risk group (favorable vs. intermediate vs. poor) and prior PD-1/PD-L1 inhibitor therapy as the stratification factors.

If there is notable imbalance in number of prior lines of anti-cancer therapy between the two treatment arms and if sample size permits, sensitivity analysis will be performed using number of prior lines of anti-cancer therapy (1 vs. 2) and IMDC prognostic risk group (favorable vs. intermediate vs. poor) and prior PD-1/PD-L1 inhibitor therapy as the stratification factors.

Similarly to the primary analysis, to prevent a stratification cell having too few events, if a stratification cell has less than 10 PFS events (two arms combined), the stratification factor of prior PD-1/PD-L1 inhibitor therapy will be removed.

Results from an unstratified analysis will also be provided.

Table 4 Censoring Rules for Progression Free Survival

Situation	Date of PFS Event or Censoring	Outcome
No baseline disease assessment	Date of randomization	Censored
No post-baseline assessments and no death	Date of randomization	Censored
No progression and no death	Date of last evaluable tumor assessment	Censored
Additional cancer therapy prior to documentation of disease progression or death	Date of last evaluable tumor assessment before initiation of additional cancer therapy	Censored
Documented RECIST v1.1 progression or death between scheduled visits [1]	First date of documented progression or death	PFS event
RECIST v1.1 progression or death documented to occur following missing at least 2 consecutive scheduled radiographic disease assessments (including missing assessments, insufficient data for the assessments, and an overall response of non-evaluable) following previous evaluable radiologic tumor assessment	Date of last tumor assessment with no documented progression [2]	Censored [2]

RECIST v1.1, Response Evaluation Criteria in Solid Tumors Version 1.1

[1] RECIST v1.1 progression or death can occur either on treatment or during the survival follow up period.

[2] Sensitivity analysis with status of PFS event at the first date of documented progression or death.

PFS (month) = (date of event / censor – date of randomization + 1) / 30.4375

6.5.2 Secondary Analyses

6.5.2.1 Overall Survival

Overall Survival (OS) is defined as the time from randomization to death from any cause. Patients who are not reported as having died at the date of analysis will be censored at the date when they were last known to be alive. Patients who do not have post-baseline information will be censored at the date of randomization.

OS will be analyzed with the same methodologies as PFS. Treatment comparisons will be based on the stratified log-rank test in the ITT population. The null and alternative hypotheses can be phrased in terms of the survival functions $S_{OS\_CB\text{-}Cabo}(t)$ and $S_{OS\_Pbo\text{-}Cabo}(t)$ in arm CB-Cabo and arm Pbo-Cabo, respectively:

$$H_0: S_{OS\_CB\text{-}Cabo\ (t)} = S_{OS\_Pbo\text{-}Cabo\ (t)} \text{ versus } H_1: S_{OS\_CB\text{-}Cabo\ (t)} \neq S_{OS\_Pbo\text{-}Cabo\ (t)}$$

OS will be tested with a total of two-sided alpha of 0.05 if PFS crosses the efficacy boundary. P-value boundaries of OS are provided in Table 3.

As a sensitivity analysis, the above analyses will be performed using the number of prior TKI cancer therapy (0 vs. ≥ 1) and IMDC prognostic risk group (favorable vs. intermediate vs. poor) and prior PD-1/PD-L1 inhibitor therapy as the stratification factors.

If there is notable imbalance in number of prior lines of anti-cancer therapy between the two treatment arms and if sample size permits, sensitivity analysis will be performed using number of prior lines of anti-cancer therapy (1 vs. 2) and IMDC prognostic risk group (favorable vs. intermediate vs. poor) and prior PD-1/PD-L1 inhibitor therapy as the stratification factors.

Similarly to the primary analysis, to prevent a stratification cell having too few events, if a stratification cell has less than 10 PFS events (two arms combined), the stratification factor of prior PD-1/PD-L1 inhibitor therapy will be removed.

Results from an unstratified analysis will also be provided.

6.5.2.2 Investigator-assessed PFS

Investigator-assessed PFS will be defined and analyzed the same way as the primary efficacy endpoint of IRC-assessed PFS. Disease progression status and date will be determined by investigators instead of IRC.

6.5.3 Additional Analyses

6.5.3.1 Overall Response Rate

Overall response rate (ORR) is defined as the percentage of patients who had an overall response (unconfirmed). An overall response is defined as either complete response (CR) or partial response (PR), as assessed by the IRC according to RECIST v1.1. Patients not meeting these criteria, including patients without any post-baseline evaluable response assessments, will be considered non-responders. Confirmation is not required in this double-blind protocol, but for exploratory purposes ORR with confirmation by IRC will also be summarized. Two CR and PR assessments must be at least 28 days apart without non-response in between, when confirmation is reported.

The analysis population for ORR will be the ITT population. An estimate of ORR and its 95% CI will be calculated with the Clopper-Pearson method for each treatment arm. The difference in ORRs between the two treatment arms and its 95% CI will be computed and compared by Fisher's exact test. The ORR will also be compared between the two arms using the stratified Cochran-Mantel-Haenszel test, stratified by the same factors used in the primary PFS and OS analyses (see Section 6.5.1).

6.5.3.2 Duration of Response

Duration of Response (DOR) is defined for patients achieving a PR or a CR as the time from the first documented overall response (PR or CR) to the first documented PD as assessed by the IRC according to RECIST v1.1 or death from any cause, whichever

occurs first. Patients who are alive and who have not experienced disease progression at the time of the analysis will be censored at the date of the last evaluable tumor assessment. If no tumor assessments were performed after the date of the first occurrence of the overall response (CR or PR), DOR will be censored at the date of the first occurrence of the overall response. Censoring rules for DOR will follow those in Table 4**.** The determination of DOR is based on a non-randomized subset of patients (confirmed responders only), and formal hypothesis testing will not be performed. The subset of responders are defined based on post randomization outcomes therefore the responder population is not balanced across the two treatment arms. DOR will be estimated using Kaplan Meier (KM) methodology. Comparisons between treatment arms will be made using the stratified and unstratified log-rank test for descriptive purposes only.

DOR will be reported for patients with confirmed ORR.

6.5.3.3 Disease Control Rate

Disease Control Rate (DCR) is defined as the summed percentage of patients with CR, PR or stable disease (SD) documented following randomization according to RECIST v1.1 per IRC.. DCR will be analyzed the same way as ORR.

6.5.3.4 PFS and OS Rates at Landmark Timepoints

The PFS rates (according to RECIST 1.1 criteria per IRC) at 6-month, 9-month, and 1-year timepoints are defined as the probabilities the patients are alive and progression free at those timepoints after randomization. The OS rates at 1- and 2- and 3-year landmark timepoints are defined as the probabilities that patients are alive at those time points after randomization. The PFS and OS rate at the landmark timepoints after randomization within the ITT population will be estimated for each treatment arm using Kaplan-Meier methodology, along with 95% CI calculated with the standard error derived from the Greenwood formula.

6.5.3.5 Patient Reported Outcomes

Patient-reported outcomes (PRO) for the NCCN-FACT FKSI-19 (Version 2) will be summarized by treatment group. Summary statistics (mean, SD, median, range) of the change from baseline will be provided at each time point. Mean score for each endpoint will be displayed graphically. The analysis will be performed for patients in the PRO-evaluable population, which is defined as ITT with a non-missing baseline assessment and at least one non-missing post-baseline assessment until treatment discontinuation.

With the exception of the visual analog scale (VAS) score and index score for EQ-5D-5L questionnaire, descriptive statistics summarizing the proportions of patients who reported having “no,” “slight,” “moderate,” “severe,” or “extreme/unable” problems at each time point will be reported. The VAS score and the US-specific Index score for EQ-5D-5L questionnaire will be summarized and analyzed as continuous measures. Patients without post-baseline assessments will be excluded from this analysis.

6.5.3.6 Additional Sensitivity Analyses

Missing Tumor Assessment

The impact of missing scheduled tumor assessments on PFS may be assessed depending on the number of patients who missed assessments scheduled immediately prior to the date of disease progression per RECIST v1.1. If > 5% of patients missed two or more scheduled assessments consecutively immediately prior to the date of disease progression per RECIST v1.1 in any treatment arm, the following sensitivity analyses will be performed:

Patients who missed two or more scheduled assessments immediately prior to the date of disease progression per RECIST v1.1 or the data cutoff or death will be censored at the last tumor assessment prior to the missed visits.

Statistical methodologies analogous to those used in the primary analysis of PFS as specified in Section 6.5.1 will be used for this sensitivity analysis.

Loss to Follow-up

The impact of loss to follow-up on OS may be assessed depending on the number of patients who are lost to follow-up. If > 5% of patients are lost to follow-up for OS in either treatment arm, a sensitivity analysis will be performed for the comparisons between two treatment arms in which patients who are lost to follow-up will be considered as having died on the date they are lost to follow-up.

6.5.3.7 Subgroup Analyses

The consistency of the primary PFS and secondary OS results in subgroups will be examined. The subgroups are defined by the following

1) demographics (age [<75 vs ≥ 75], sex [F vs M], and race/ethnicity)

2) baseline prognostic characteristics (KPS performance status [70-80, 90-100], IMDC prognostic risk grouping [EDC version], number of organs with metastases [1, 2, ≥3], prior nephrectomy, and prior radiotherapy, time from metastatic diagnosis [>1 and <= 1 year])

3) Prior therapies (previous anti-angiogenic therapies [yes/no], number of prior TKI [1, 2, ≥3], prior PD-1/PD-L1 [yes, no], prior nivo/ipi [yes, no], number of prior line of therapy [2, 3])

4) Metastatic Sites (lung, lymph node, bone, liver, brain, other as categories)

5) Geographic region (US/ex US)

Summaries of PFS and OS, including unstratified HRs estimated from Cox proportional hazards models and Kaplan-Meier estimates of median PFS and OS, will be produced separately for each level of the categorical variables for the comparisons between treatment arms and displayed in a forest plot (Lewis 2001).

Summaries of ORR by subgroup defined above will also be provided.

Analyses of subgroups with sample size too small may not be performed.

6.6 Pharmacokinetic Analyses

Population PK analysis of telaglenastat will be performed for the PK-evaluable population, defined as all CB-Cabo patients who have received at least one dose of telaglenastat and have at least one evaluable PK sample. Individual PK concentrations will be tabulated and summarized. Descriptive statistics will include means, medians, ranges, and SDs, as appropriate.

The impact of telaglenastat on the PK exposure of cabozantinib will be evaluated by comparing the CB-Cabo arm with that of the Pbo-Cabo arm.

Additional PK analyses will be conducted as appropriate, on the basis of the availability of data. A separate report will include these additional analyses as well as details of population PK analyses.

6.7 Safety Analyses

Unless specified otherwise, safety analyses described in this section will be conducted for the Safety population, with patients grouped according to whether they received any telaglenastat treatment.

6.7.1 Exposure of Study Medication

Study drug exposure status for each component of the study drug, which include treatment duration, total dose received (mg), number of cycles, number of patients with dose adjustment and the reasons will be summarized for each treatment arm for the Safety population.

6.7.2 Adverse Events

Verbatim description of AEs will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) preferred terms and system organ classes, and graded by the investigator according to NCI CTCAE v4.03. Treatment-emergent adverse events (TEAEs) will be summarized by mapped term, appropriate thesaurus level, NCI CTCAE grade, and treatment arm. In addition, serious adverse events, severe adverse events

(Grade ≥ 3), and adverse events leading to study drug discontinuation or interruption will be summarized accordingly. These tables will also be presented for TEAEs judged to be related (probably or possibility) to the study treatment. Multiple occurrences of the same event in a given patient will be counted once at the maximum grade. All deaths are reported as SAEs per study conduct. In order to accurately summarize the true SAE rate, all SAE summaries will not count Grade 5 events for patients who died due to progressive disease.

Per protocol and study conduct, safety follow-up is 28 days from date of last study drug administration or until initiation of a new anti-cancer therapy if earlier. For reporting purposes, "treatment emergent" is defined as new adverse events occurring on or after the first dose of study drug is administered until the clinical cutoff date, or existing events that worsened after the first dose during the study.

All listings of adverse events will also include all adverse events.

Deaths reported during the study treatment period and the follow-up period after treatment discontinuation will be listed by treatment arm.

6.7.3 Laboratory Data

Laboratory data will be summarized by treatment arm at each assessment time point. Values outside the normal ranges will be summarized by treatment arm. In addition, selected laboratory data will be summarized by treatment arm and NCI CTCAE v4.0 grade at each assessment time point.

6.7.4 Vital Signs

Changes in selected vital signs will be summarized by treatment arm and by change over time, which includes change from baseline at each visit where parameters are scheduled to be collected. Baseline is defined as the measurement obtained on Cycle 1, Day 1 before the first dose of study drug is administered.

6.8 Missing Data

See Section 6.5 for methods for handling missing data for primary and secondary efficacy endpoints. Missing data will not be imputed for safety analyses except for the purpose of identification of TEAEs and prior or concomitant medications with missing start or end date. Imputation rules for the partial or missing dates will be pre-specified in a separate document. All available data will be presented on the data listings as collected (no imputation).

6.9 Analyses Related to COVID-19

This section describes the pre-specified analyses to assess the potential impact of the COVID-19 pandemic on the study conduct and the evaluation of efficacy and safety of the drug.

Two populations defined:

COVID-19 impacted: Any patients with important protocol deviations due to COVID-19 including missed important visits, missed data and changes in methods of important assessments or those who are infected with COVID-19

COVID-19 infected: any patient with a reported infection (a subgroup of COVID-19 impacted)

6.9.1 General

Important protocol deviations due to COVID-19 including missed important visits, missed data and changes in methods of important assessments will be summarized by treatment groups.

Number and percent of patients who are COVID-19 infected and impacted will be summarized by treatment groups. The demographic and baseline disease characteristics and extent of exposure to study drug will be summarized for these patients.

Potential additional analyses with COVID-19 infected and COVID-19 impacted may be done for any of the endpoints including patient reported outcomes etc.

Listing of all patients that are COVID-19 impacted and infected will be reported.

6.9.2 Efficacy

6.9.2.1 Progression Free Survival

Two kinds of events due to COVID-19 can impact the PFS endpoint: deaths and missing tumor assessments.

The first one is missing tumor assessment due to COVID-19 (COVID-19 impacted). Since primary analysis will handle missing tumor assessment due to cause including COVID-19 (described in Table 4), this scenario is covered by the primary analysis missing tumor assessment handling.

The second one is if there are a total of 10 or more patients who die due to COVID-19 (COVID-19 infected) without a documented disease progression or starting a new anti-cancer therapy or missing at least 2 consecutive scheduled radiographic disease assessments for reasons not related to COVID-19, the following sensitivity analyses will be performed:

a. These patient will be considered as censored at the date of death due to COVID-19. The primary analyses described in section 6.5.1 (log-rank test and Cox proportional hazard model) will be repeated using this adjusted censoring rule.

b. Inverse probability of censoring weighting (IPCW) approach will be used to adjust for potential bias of the planned primary analysis introduced by COVID-19 death. With this approach, COVID-19 death will be modeled using a Cox model with treatment group and factors specified in section 6.5.3.7 as covariates. Hazard for COVID-19 death will be estimated for each subject based on which weights will be obtained. The analyses described in section 6.5.1 (Cox regression model and KM estimates) will be performed using these weights.

6.9.2.2 Overall Survival

If there are a total of 10 or more patients who die due to COVID-19, similar sensitivity analyses described for the PFS (a and b) above will be performed for the OS endpoint.

6.9.3 Safety

Treatment-emergent AEs and SAEs due to COVID-19 infection will be summarized by treatment groups, preferred terms, and severities. All AEs for subjects diagnosed with a COVID-19 infection will be summarized by treatment groups, preferred terms, severities and the relatedness to the study drug.

7 REFERENCES

1. Brookmeyer, R. and Crowley, J. (1982) A confidence interval for the median survival time. Biometrics, 38, 29-41. doi:10.2307/2530286

2. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical principles for clinical trials (E9). International Conference on Harmonization; 1998.

3. M2 eCTD: Electronic Common Technical Document Specification Appendix 7, provided by the International Conference on Harmonization. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073240.pdf

4. Data Standards: Position on Use of SI Units for Lab Tests. U.S Food and Drug Administration; 25 October 2013. Available from: http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm372553.htm

5. U.S. Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE), Version 4.03. 14 June 2010. Available from: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf.

6. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322:1479-80.

7. Heng D Stukalin I, Wells C, Donskov F, Rini BI, Lee JL, et al, Third-line therapy in metastatic renal cell carcinoma (mRCC): Results from the International mRCC Database Consortium (IMDC). J Clin Oncol 2015; 30:suppl; abstr e15578.