Official Title of Study:

A Phase 3, Randomized, Open-Label Study of Nivolumab Combined with Cobazantinib Versus Sunitinib in Participants with Previously Untreated Advanced or Metastatic Renal Cell Carcinoma

NCT Number: NCT03141177

Document Date (Date in which document was last revised): Dec 11, 2019

STATISTICAL ANALYSIS PLAN

A PHASE 3, RANDOMIZED, OPEN-LABEL STUDY OF NIVOLUMAB COMBINED WITH CABOZANTINIB VERSUS SUNITINIB IN PARTICIPANTS WITH PREVIOUSLY UNTREATED ADVANCED OR METASTATIC RENAL CELL CARCINOMA

PROTOCOL CA2099ER

VERSION # 2.0

DATE: 11-DEC-2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

[redacted]

Research Hypothesis:

Treatment with nivolumab combined with cabozantinib (doublet regimen) will demonstrate an improvement in PFS per BICR compared to sunitinib monotherapy in subjects with previously untreated mRCC.

Schedule of Analyses:

This study will be monitored by an independent Data Monitoring Committee (DMC). Details are specified in the DMC charter.

The PFS analysis will occur after approximately 9-10 months minimum follow-up on all randomized subjects, which will be triggered by approximately 350 events from Arm A (nivolumab combined with cabozantinib; doublet regimen) and Arm C (sunitinib).

Two interim analyses of OS are planned. The first interim analysis is planned at the time of final PFS analysis and expected after observing 165 deaths among the randomized subjects in Arm A and Arm C (65% of the targeted OS events for final analysis). The second interim analysis is expected after observing 211 deaths among the randomized subjects in Arm A and Arm C (83% of targeted OS events needed for final analysis. The final analysis of OS is expected after observing 211 deaths among the randomized subjects in Arm A and Arm C.

The final PFS analysis will not occur prior to these conditions being met:

- at least 8 months minimum follow-up on all randomized subjects;
- at least 283 PFS events, which provide at least 90% power to detect a HR of 0.68 for PFS of Arm A versus Arm C; and
- at least 149 OS events, which provide 66% power if the observed HR for OS was 0.60.

The expected PFS analysis will occur at approximately 29 months from FPFV. The second interim and final analyses of OS are expected to occur approximately 34 months and 40 months from FPFV.

Secondary endpoints (including both efficacy endpoints OS and ORR) will be analyzed at the time of the final analysis of PFS based on a hierarchical testing strategy. In the event that the interim analysis for superiority of overall survival is positive, final (CSR) analyses will be performed prior to achieving 254 deaths; additional details can be found in section 7.5.7.

2 STUDY DESCRIPTION

Implementation of CA2099ER Global Revised Protocol 01 stops further randomization into Arm B (nivolumab + ipilimumab combined with cabozantinib). Subjects previously randomized to Arm B continue with Arm B treatment and continue with Arm B clinical planned events, per protocol.

This is an open label, randomized trial of nivolumab combined with cabozantinib (doublet regimen) versus sunitinib in subjects with previously untreated (first line) advanced or metastatic RCC. Subjects will be randomized between Arm A and Arm C in a 1:1 ratio with approximately 638 subjects (319 per arm) capped at approximate 25% to represent the normal frequency of favorable risk group in mRCC. The rest of the randomized subjects will provide approximately 478 intermediate/poor risk randomized subjects (239 per arm). Subjects will be stratified at the time of randomization by IMDC prognostic score (0 [favorable risk] versus 1-2 [intermediate risk] versus 3-6 [poor risk]), PD-L1 tumor expression (≥ 1% versus < 1% or indeterminate), and region (US/Canada/Western Europe/Northern Europe versus rest of the world [ROW]).

The subject is randomly assigned to 1 of the 2 treatment arms as noted in the study schematic below.

- Arm A (Doublet): Nivolumab 240 mg flat dose IV Q2W + Cabozantinib 40 mg PO QD
  - Nivolumab to be continued until disease progression, unacceptable toxicity, or a maximum treatment of 2 years from the first dose in Cycle 1.
  - Cabozantinib to be continued until disease progression or unacceptable toxicity.
- Arm C: Sunitinib 50 mg PO QD for 4 weeks, followed by 2 weeks off, per cycle. Cycles to be continued until progression or unacceptable toxicity.

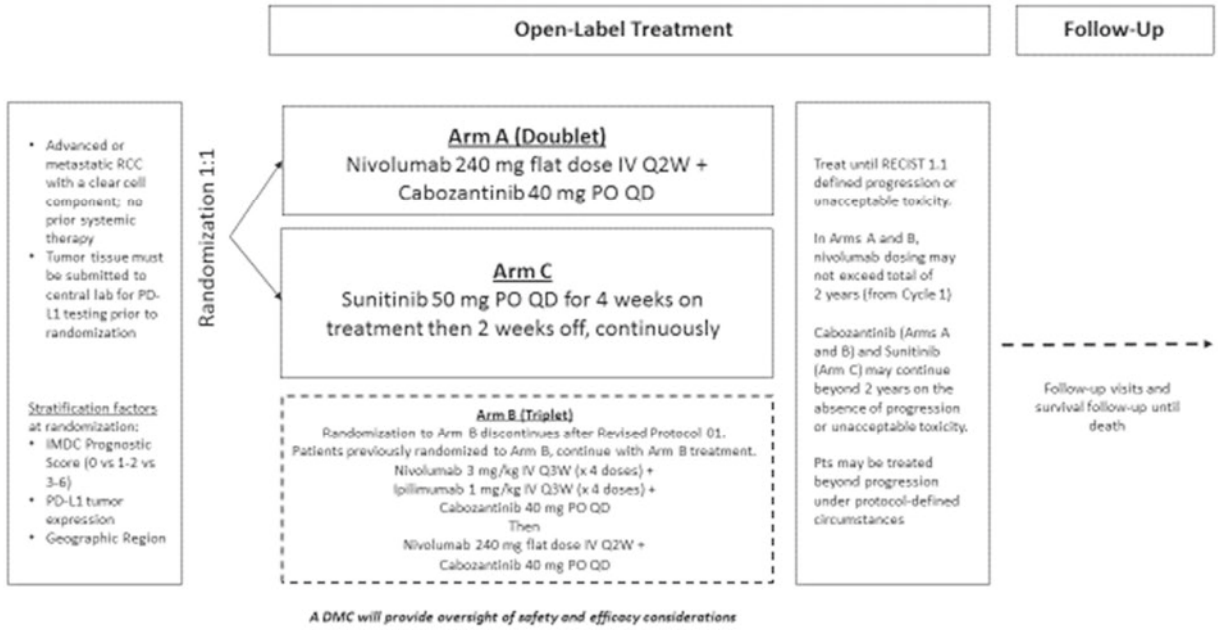
  Note - Randomization to Arm B stops with implementation of approved CA2099ER Global Revised Protocol 01. Treatment B continues only for subjects randomized to Arm B prior to implementation of Global Revised Protocol 01.

  - Arm B (Triplet): Nivolumab 3mg/kg IV + Ipilimumab 1 mg/kg IV, both Q3W x 4 doses + Cabozantinib 40 mg PO QD
    - Then Nivolumab 240 mg flat dose IV Q2W + Cabozantinib 40 mg PO QD.
    - Nivolumab to be continued until disease progression, unacceptable toxicity, or a maximum of 2 years from the first dose in Cycle 1.
    - Cabozantinib to be continued until progression or unacceptable toxicity.

Once randomized subjects in Arm A will continue nivolumab until progression, unacceptable toxicity, withdrawal of consent, or a maximum of 2 years from the first dose in Cycle 1, whichever occurs first. Cabozantinib (Arm A) may be continued until progression, unacceptable toxicity, or withdrawal of consent, whichever occurs first, and may extend beyond 2 years from the first dose in Cycle 1.

The study design schematic is presented in the figure below.

Figure 2-1: Study Design Schematic



2.1 Treatment Assignment

After the subject's initial eligibility is established and informed consent has been obtained, the subject must be enrolled into the study by accessing an Interactive Response Technologies web-based system (IRT) to obtain the subject number. All subjects will be centrally randomized using an Interactive Response Technology (IRT). Before the study is initiated, each user will receive log in information and directions on how to access the IRT.

Every subject that signs the informed consent form must be assigned a subject number in IRT. The investigator or designee will register the subject for enrollment by following the enrollment procedures established by BMS. The following information is required for enrollment:

- Date that informed consent was obtained
- Date of birth
- Sex at birth.

Once enrolled in IRT, enrolled subjects that have met all eligibility criteria will be ready to be randomized through the IRT. The following information is required for subject randomization:

- Subject number
- Date of birth
- IMDC Prognostic Score (0 versus 1-2 versus 3-6)
- Region (US/Canada/W Europe/N Europe versus ROW)

- PD-L1 tumor expression (≥ 1% versus < 1% or indeterminate)

Subjects meeting all eligibility criteria will be randomized in a 1:1 ratio to Arm A (nivolumab combined with cabozantinib) or Arm C (sunitinib), stratified by the following factors:

- IMDC Prognostic Score: 0 versus 1-2 versus 3-6
- Region: US/Canada/W Europe/N Europe versus ROW
- PD-L1 tumor expression: ≥ 1% versus < 1% or indeterminate

The randomization procedures will be carried out via permuted blocks within each stratum.

2.2 Blinding and Unblinding

This is an open label study.

2.3 Protocol Amendments

Table 2.3-1: Protocol Amendments

Document	Date of Issue	Summary of Change
Revised Protocol 02	06-MAY-2019	Major Changes: • Revised protocol 02 adjusts the timing of the PFS and OS interim analyses with modified hypothesized OS hazard ratio (HR). The number of randomized subjects is increased. • The interim analysis for ORR is removed, resulting in revised overall alpha for PFS and OS endpoints. • No change in eligibility or study procedure. • Clinical data for nivolumuab + ipilimumab in RCC has been updated. Other changes include more detail on PRO measures and updates to align with BMS standards for the nivolumab program.
Revised Protocol 01	18-DEC-2017	Primary revisions: (i) To stop enrollment into Arm B (nivolumab, ipilimumab and cabozantinib triplet) and (ii) to include favorable risk subjects (capped at 25%) in the primary data analysis. Secondary items include: (i) to add a Data Monitoring Committee review after 30 subjects are treated for 6 weeks, (ii) to adjust, clarify and add exclusion criteria, (iii) to add treatment restrictions, (iv) to clarify criteria associated with hemorrhage with regard to resuming treatment, (v) to specify an additional precaution when sunitinib dosing is resumed, and (vi) to apply newly updated Sponsor standards for nivolumab clinical protocols. Tertiary items include (i) incorporation of Administrative Letter 01 and (ii) correction of typographical and grammatical errors.

2.4 Data Monitoring Committee

An independent Data Monitoring Committee (DMC) has been established to provide oversight of safety and efficacy considerations, study conduct, and risk-benefit ratio. Following review, the DMC will recommend continuation, modification, or discontinuation of this study based on reported safety and efficacy data. Details of DMC responsibilities and procedures are specified in the DMC charter. Representatives of the Sponsor will serve only as coordinators of the committee, without having full member responsibilities or privileges. In addition, the Sponsor will independently review safety data in a blinded manner during the conduct of this trial to ensure that any safety issues are identified and addressed.

The DMC will conduct the first review of the safety data after at least 30 subjects are treated and followed for at least 6 weeks. The DMC will conduct its second review of the safety data after at least 75 subjects are treated and followed for at least 6 weeks. The DMC will conduct its review of the safety data focusing on the initial approximately 12 Japanese subjects (6 per arm) treated and followed for at least 4 weeks. The DMC will then review safety and the available efficacy data pertaining to primary endpoint to evaluate safety in the context of benefit, every six months thereafter.

The DMC will also review the formal final analysis of PFS (as per BICR) and first interim analysis of superiority of OS scheduled at around 29 months from FPFV. A second interim analysis of OS will be at around 34 months from FPFV. BMS will remain blinded to OS interim analyses unless DMC decides to disclose the formal interim analysis to BMS.

Details of the interim analyses can be found in section 7.5.7.

2.5 Blinded Independet Central Review

A blinded independent central review (BICR) committee has been established to provide an independent imaging review of images obtained in subjects participating in this study. Details of BICR responsibilities and processes may be found in the BICR Charter. The BICR determined PFS and ORR endpoints will be utilized as a part of primary and secondary efficacy analyses.

3 OBJECTIVES

3.1 Primary

- To compare PFS per BICR of nivolumab combined with cabozantinib (Arm A: doublet) with sunitinib (Arm C) in all randomized subjects.

3.2 Secondary

- To compare overall survival (OS) of Arm A with Arm C in all randomized subjects.
- To evaluate the objective response rate (ORR) of Arm A with Arm C per BICR in all randomized subjects.
- To assess overall safety and tolerability in all treated subjects.

4 ENDPOINTS

4.1 Primary Endpoints

Progression-free survival (PFS) is the primary endpoint. Two definitions are used for the analysis of PFS. The primary definition accounts for subsequent therapy by censoring at the last evaluable tumor assessment on or prior to the date of subsequent therapy. The secondary definition is irrespective of subsequent therapy and does not account for subsequent therapy.

Clinical deterioration in the absence of unequivocal evidence of progression (per RECIST v1.1 criteria) is not considered progression for purposes of determining PFS.

PFS rate at time *T* is defined as the probability that a subject has not progressed and is alive at time *T* following randomization. PFS rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) are defined as the probability that a subject has not progressed and is alive at time *T* following randomization.

The first on-study tumor assessment is scheduled to be conducted at 12 weeks (± 1 week) following randomization. Subsequent tumor assessments are scheduled every 6 weeks (± 1 week) up to week 60, then every 12 weeks (± 2 weeks) until disease progression.

4.1.1 Primary Definition of Progression-Free Survival (Accounting for Subsequent Therapy)

The primary definition of PFS (PFS truncated at subsequent therapy, which includes anti-cancer therapy, tumor directed radiotherapy, or tumor directed surgery) is defined as the time between the date of randomization and the date of first documented tumor progression, based on BICR assessments (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first.

Subjects who die without a reported progression will be considered to have progressed on the date of their death. The following censoring rules will be applied for the primary definition of PFS:

- Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment.
- Subjects who did not have any on study tumor assessments and did not die will be censored on their date of randomization.

- Subjects who receive subsequent anti-cancer therapy prior to documented progression will be censored at the date of the last evaluable tumor assessment conducted on or prior to the date of initiation of the subsequent anti-cancer therapy.
- Subjects who did not have a documented progression and received subsequent anti-cancer therapy will be censored at the date of the last evaluable tumor assessment conducted on or prior to the initiation of the subsequent anti-cancer therapy.

Censoring rules for the primary definition of PFS (PFS truncated at subsequent therapy) are presented as follows and in Table 4.1.1-1.

Figure 4.1.1-1: PFS Primary Definition

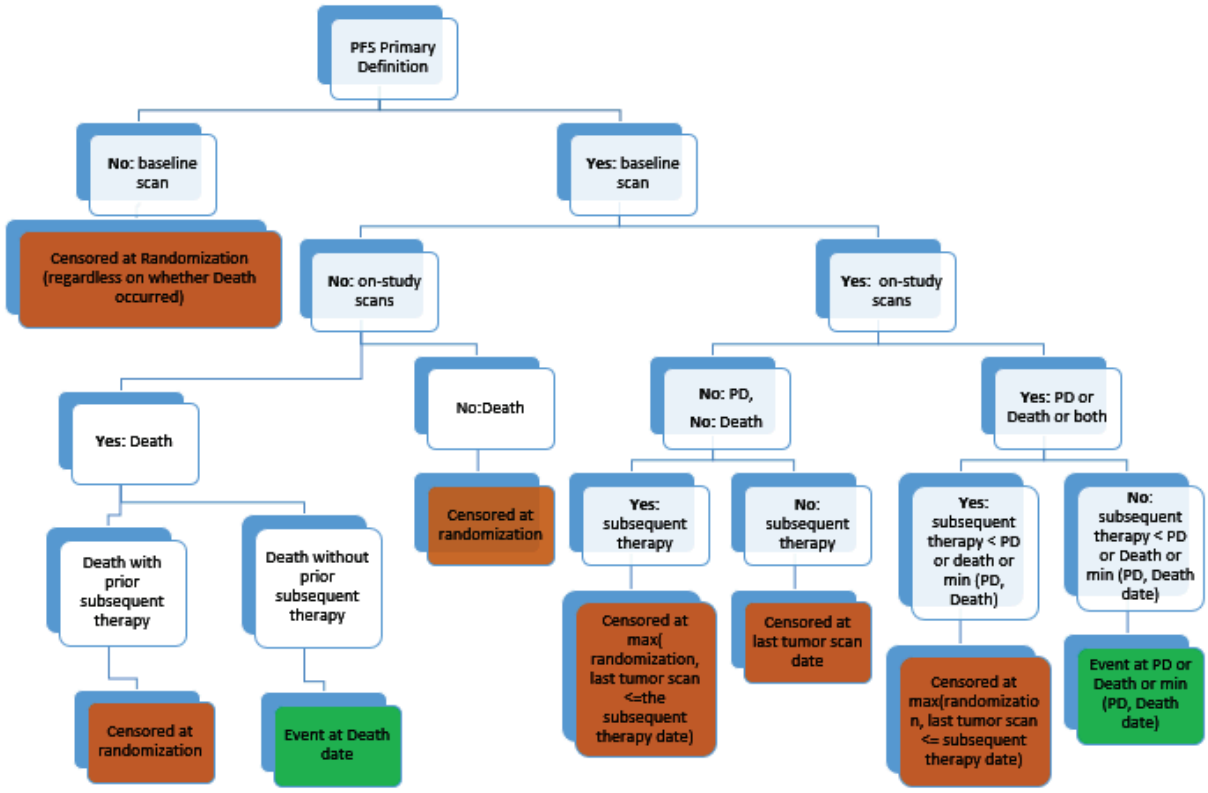


Table 4.1.1-1: Censoring Scheme Used in Primary Definition of PFS

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments*	Date of randomization	Censored
No on study tumor assessments and no death*	Date of randomization	Censored
Subsequent anti-cancer therapy started without death or progression per RECIST v1.1 reported prior or on the same day	Date of last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored

Table 4.1.1-1: Censoring Scheme Used in Primary Definition of PFS

Situation	Date of Progression or Censoring	Outcome
Documented progression per RECIST v1.1 and no new anti-cancer started before	Date of the first documented progression per RECIST v1.1 (excludes clinical progression)	Progressed
No progression and no death, and no new anti-cancer therapy started	Date of last evaluable tumor assessment	Censored
Death without progression per RECIST v1.1 and no new anti-cancer started before	Date of death	Progressed

* Tumor assessments and death if any, occurring after start of subsequent anti-cancer therapy are not considered.

4.1.2 Secondary Definition of Progression Free Survival (Irrespective of Subsequent Therapy)

The secondary definition of PFS (ITT definition) is defined as the time between the date of randomization and the date of first documented tumor progression, based on BICR assessments (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first.

Subjects who die without a reported progression will be considered to have progressed on the date of their death. The following censoring rules will be applied for the secondary definition of PFS:

- Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment.
- Subjects who did not have any on study tumor assessments and did not die will be censored on their date of randomization.

Censoring rules for the secondary definition of PFS (ITT definition) are presented as follows and in Table 4.1.1-1.

Figure 4.1.2-1: PFS Secondary Definition

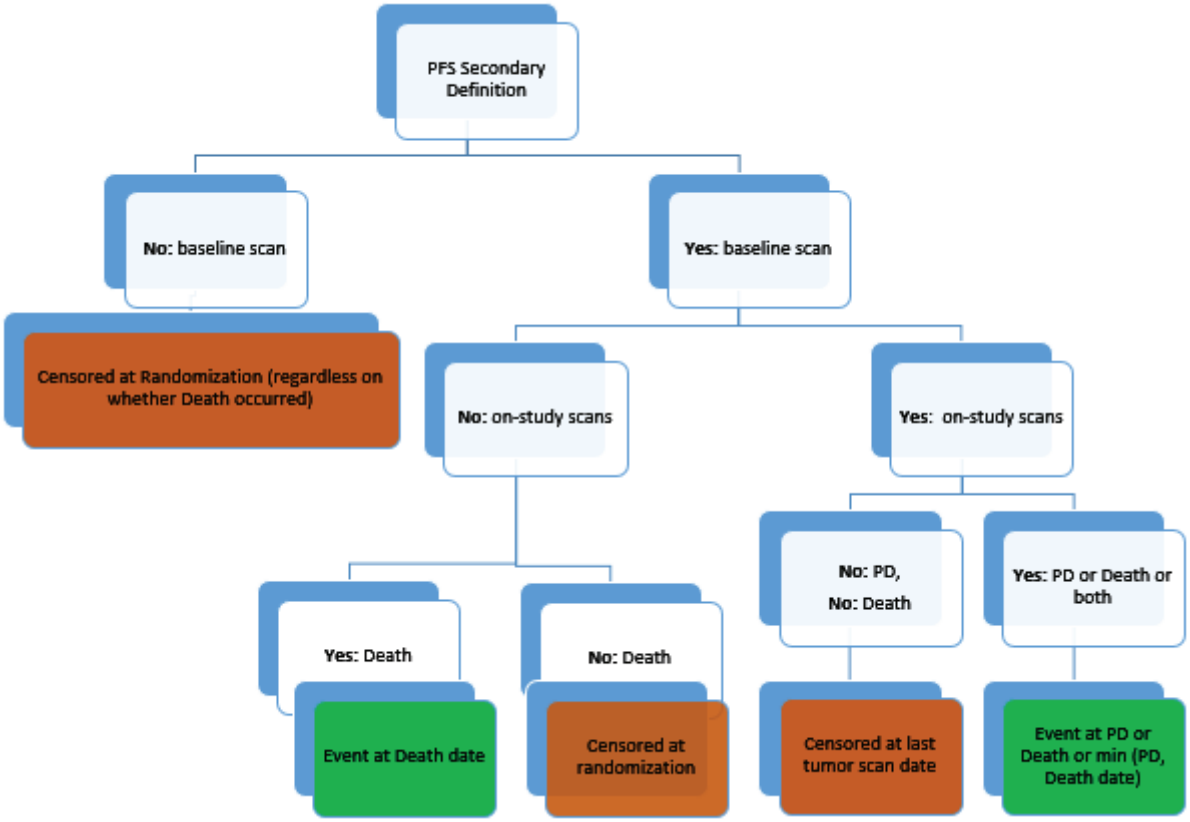


Table 4.1.2-1: Censoring Scheme for Secondary Definition of PFS

Situation	Date of Progression of Censoring	Outcome
No baseline tumor assessment	Date of randomization	Censored
No on-study tumor assessments and no death	Date of randomization	Censored
Documented progression per RECIST v1.1	Date of first documented progression per RECIST v1.1 criteria (excludes clinical progression)	Progressed
No progression and no death	Date of last evaluable tumor assessment	Censored
Death without progression per RECIST v1.1	Date of death	Progressed

Note that the secondary definition will only be used as supportive analysis.

4.2 Secondary Endpoints

4.2.1 Overall Survival

Overall survival (OS) is defined as the time from randomization to the date of death from any cause. For subjects that are alive, their survival time will be censored at the date of last contact date (or "last known alive date"). Overall survival will be censored at the date of randomization for subjects who were randomized but had no follow-up.

Follow-up visit #1 (FU1) should occur 30 days from the last dose and follow-up visit #2 (FU2) occurs approximately 100 days from last dose of study drug. After FU2, survival follow-up will be conducted every 3 months.

4.2.2 Objective Response Rate

Objective Response Rate (ORR) is defined as the number of randomized subjects who achieve a best response of confirmed complete response (CR) or confirmed partial response (PR) based on BICR assessments (using RECIST v1.1 criteria) divided by the number of all randomized subjects. Best Overall Response (BOR) is defined as the best response, as determined by the BICR, recorded between the date of randomization and the date of objectively documented progression per RECIST v1.1 criteria or the date of subsequent therapy (including tumor-directed radiotherapy and tumor-directed surgery), whichever occurs first. For subjects without documented progression or subsequent therapy, all available response designations will contribute to the BOR determination. Confirmation of response is required at least 4 weeks after the initial response.

4.2.2.1 Time to Response

Time to Response (TTR) is defined as the time from randomization to the date of the first confirmed documented response (CR or PR), as assessed by the BICR. TTR will be evaluated for responders (confirmed CR or PR) only.

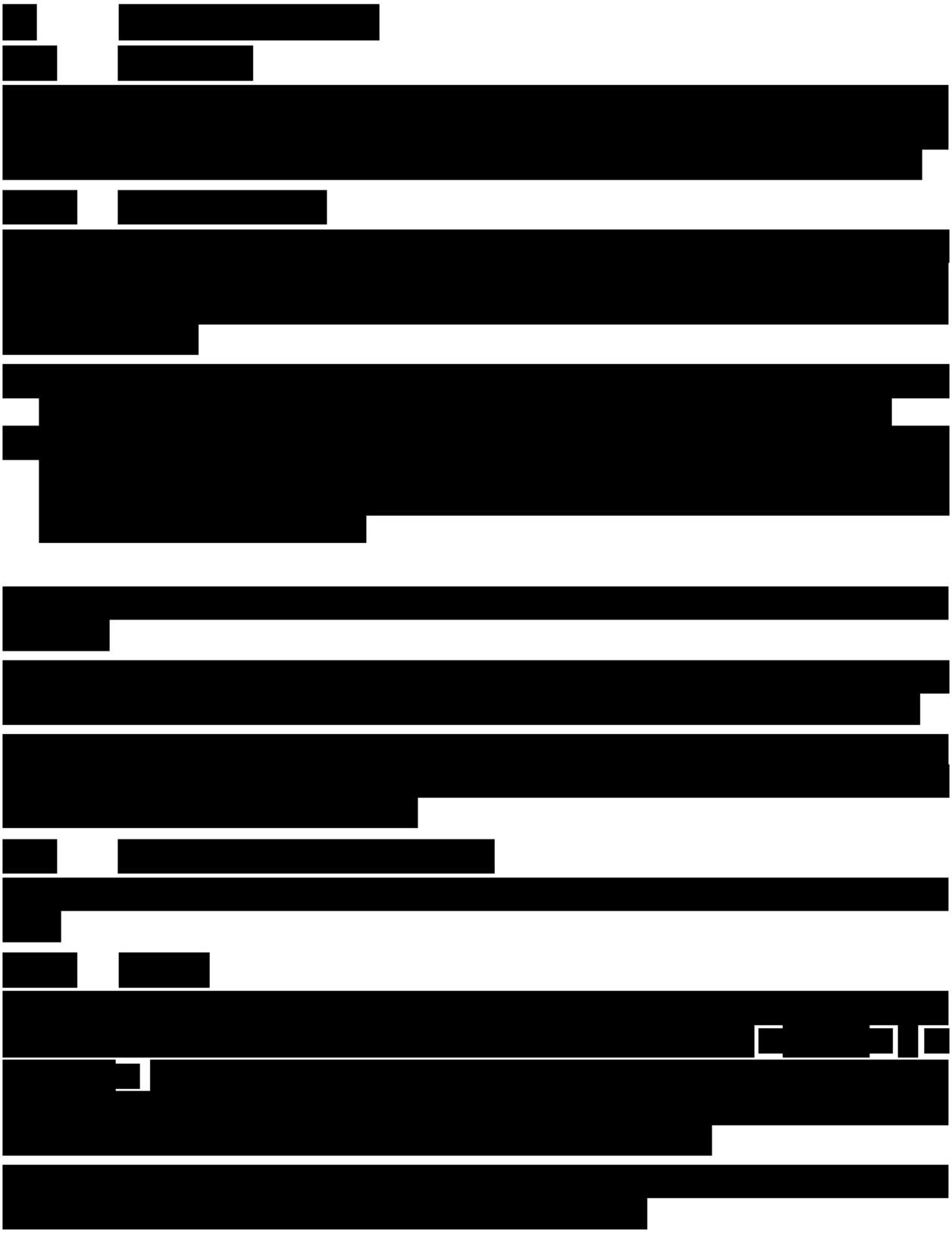
4.2.2.2 Duration of Response

Duration of Response (DOR) is defined as the time between the date of first confirmed documented response (CR or PR) to the date of the first documented tumor progression as determined by the BICR (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first. Subjects who start subsequent therapy without a prior reported progression will be censored at the last evaluable tumor assessments prior to initiation of the subsequent anti-cancer therapy. Subjects who die without a reported prior progression will be considered to have progressed on the date of their death. Subjects who neither progress nor die, DOR will be censored on the date of their last evaluable tumor assessment. DOR will be evaluated for responders (confirmed CR or PR) only.

4.3 Safety Endpoints

The assessment of safety will be based on the incidence of adverse events (AEs), serious adverse events (SAEs), adverse events leading to discontinuation, adverse events leading to dose modification, select adverse events (select AEs) for EU/ROW Submissions, immune-mediated AEs (IMAEs) for US Submission, other events of special interest (OEOSI), and deaths. The use

of immune modulating concomitant medication will be also summarized. In addition clinical laboratory tests, and immunogenicity (i.e. development of anti-drug antibody) will be analyzed.



5 SAMPLE SIZE AND POWER

The sample size calculations of this study summarized below are based on the target randomized subjects in Arm A and Arm C only. That is, the total randomized subjects will be higher than the target randomized subjects due to randomized subjects into Arm B prior to implementation of CA2099ER Global Revised Protocol 01, which stopped further randomization into Arm B.

The sample size of this study accounts for the primary endpoint of progression-free survival (PFS) per BICR in all randomized subjects. Assuming a 25% screen failure rate, it is expected that approximately 850 subjects will need to be enrolled in order to randomize 638 subjects (319 per arm) in a 1:1 ratio. To represent the normal frequency of the favorable risk group in mRCC, the favorable risk subjects are capped at approximate 25%; thus, at most 212 favorable risk subjects (106 per arm) will be enrolled to randomize 160 favorable risk subjects in a 1:1 ratio. The rest of the enrolled participants will provide approximately 478 intermediate/poor risk randomized subjects (239 per each arm).

The overall alpha for this study is 0.05 (two-sided). PFS will be evaluated for treatment effect at an alpha of 0.05 (two-sided), with at least 95% power. No interim analysis of PFS is planned. OS will be evaluated for treatment effect at an alpha level of 0.05 (two-sided) with 80% power, accounting for two formal interim analyses to assess efficacy.

Sample Size Justification for Primary PFS Endpoint

The primary endpoint of PFS per BICR of Arm A versus Arm C analysis conducted on all randomized subjects. The PFS analysis will occur after approximately 9-10 months minimum

follow-up on all randomized subjects, which will be triggered by approximately 350 events from Arm A and Arm C. The 350 PFS events provide at least 95% power to detect a HR of 0.68 for PFS of Arm A versus Arm C with a type I error of 0.05 (two-sided). The HR of 0.68 corresponds to a 47% increase in the median PFS, assuming a median PFS of 18.2 months for Arm A and 12.4 months for Arm C. It is projected that an observed HR of 0.811 or less, which corresponds to a 2.89 month or greater improvement in median PFS (12.4 versus 15.3 months), would result in a statistically significant improvement in PFS for the Arm A versus Arm C comparison.

If the formal analysis of PFS among all randomized subjects is statistically significant, the formal interim analysis of OS among all randomized subjects will be tested, as per hierarchical testing procedure. If the formal analysis of OS (interim or final, whichever occurs first) among all randomized subjects is statistically significant, then formal analysis of ORR among all randomized subjects will be tested, as per hierarchical testing procedure. The formal ORR analyses specified in the SAP supersede those specified in the protocol.

Sample Size Computation for Secondary OS Endpoint

The secondary endpoint of OS in all randomized subjects specifies the comparison of Arm A versus Arm C. Among all randomized subjects, approximately 254 events (ie, deaths) in Arm A and Arm C provides at least 80% power to detect a HR of 0.70 for OS of Arm A and Arm C with an overall type 1 error of 0.05 (two-sided) for each test. The HR of 0.70 corresponds to a 43% increase in the median OS, assuming a median OS of 47.1 months for Arm A and 33 months for Arm C.

Two formal interim analyses of OS are planned for this study. The first interim analysis is planned at the time of final PFS analysis and it is expected to observe 165 OS events (65% of the targeted OS events for final analysis) and the second interim analysis is planned to occur after observing approximately 211 events (83% of targeted OS events needed for final analysis). The stopping boundaries at interim and final analyses will be derived based on the number of deaths using O'Brien and Fleming α spending function. For example, with 165, 211, and 254 observed events in Arm A and Arm C at the first interim, second interim, and final analyses, the respective stopping boundaries would be α=0.011 (two-sided), α=0.025 (two-sided), and α=0.041 (two-sided). If the first interim analysis is performed exactly at 165 deaths, it is projected that an observed HR of 0.673 or less, which corresponds to a 16.0 month or greater improvement in median OS (33 versus 49 months), would result in a statistically significant improvement in OS for the Arm A versus Arm C comparison. At the second interim analysis with 211 deaths, it is projected that an observed HR of 0.734 or less, which corresponds to a 12.0 month or greater improvement in median OS (33 versus 45 months), would result in a statistically significant improvement in OS for the Arm A versus Arm C comparison. At the time of final OS analysis when there are 254 deaths, it is projected that an observed HR of 0.774 or less, which corresponds to a 9.6 month or greater improvement in median OS (33 versus 42.6 months), would result in a statistically significant improvement in OS for the Arm A versus Arm C comparison.

Assuming a constant accrual rate (an average rate of 3 subjects/month in the first 4 months, afterwards an average rate of 42 subjects/month), the accrual will take approximately 19 months. The final PFS analysis will not occur prior to these conditions being met:

- at least 8 months minimum follow-up on all randomized subjects;
- at least 283 PFS events, which provide at least 90% power to detect a HR of 0.68 for PFS of Arm A versus Arm C; and
- at least 149 OS events, which provide 66% power if the observed HR for OS was 0.60.
  (Note that if the analysis of first interim analysis OS takes place with 149 OS events, the alpha spending for the OS comparison would be 0.007 with a critical HR=0.643.)

This expected PFS analysis will occur at approximately 29 months from FPFV. The second interim and final analyses of OS are expected to occur approximately 34 months and 40 months from FPFV, respectively. Table 5-1 summarizes the results of these calculations.

Table 5-1: Summary of Sample Size Parameters and Schedule of Analyses

Primary/Secondary Endpoints	PFS (Primary)	OS (Secondary)
Primary analysis population	All Randomized Subjects	
Accrual rate per month for all randomized population	3 subjects/month in the first 4 months, afterwards an average rate of 42 subjects/month	
Power	95%	80%
Alpha	0.05 2-sided	0.05 2-sided (0.011 at IA1, 0.025 at IA2, 0.041 at FA)
Hypothesized median control vs exp (months)	12.4 vs 18.2	33 vs 47.1
Hypothesized hazard ratio	0.68	0.70
Critical hazard ratio (observed hazard ratio at which a statistically significant difference would be observed) / Difference in median (months) Corresponding to a minimal clinically significant effect size (FA)	0.811 / 2.89	0.774 / 9.6
Critical HR at interim analysis-1(IA1) /effect size	N/A	0.673 / 16.0
Expected number of event for IA1 (percentage of target events)	N/A	165 (65%)
Timing of IA1 from FPFV (months)	N/A	29
Critical HR at interim analysis-2(IA2) /effect size	N/A	0.734 / 12.0
Expected number of event for IA2 (percentage of target events)	N/A	211 (83%)
Timing of IA2 from FPFV (months)	N/A	34
Accrual duration (months)	19	19

Table 5-1: Summary of Sample Size Parameters and Schedule of Analyses

Primary/Secondary Endpoints	PFS (Primary)	OS (Secondary)
Timing of final analysis (FA) from FPFV (months)	29	40
Sample size	638	638
Target number of events (Event Goal)	350	254

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

- Baseline period:
  - Baseline evaluations or events will be defined as evaluations or events that occur before the date and time of the first dose of study treatment. Evaluations (laboratory tests, pulse oximetry and vital signs) on the same date and time of the first dose of study treatment will be considered as baseline evaluations. Events (AEs) on the same date and time of the first dose of study treatment will not be considered as pre-treatment events.
  - In cases where the time (onset time of event or evaluation time and dosing time) is missing or not collected, the following definitions will apply:
    - Pre-treatment AEs will be defined as AEs with an onset date prior to but not including the day of the first dose of study treatment;
    - Baseline evaluations (laboratory tests, pulse oximetry and vital signs) will be defined as evaluations with a date on or prior to the day of first dose of study treatment.
  - If there are multiple valid assessments on or prior to the first dose of study treatment:
    - For laboratory tests, the latest non missing labs value on or before first dose date (and time if collected) will be used as the baseline in the analyses. For 'LIPASE' and 'GLUCOSE', for treated subjects only, the last predose assessment with non-missing toxicity grade will be considered as baseline. If multiple assessments exist with the same collection date (and time if collected) and entry date and time, then the first observation is used as baseline.

[redacted]

[redacted]

- Post baseline period:
  - On-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of study treatment (or with an onset date on or after the day of first dose of study treatment if time is not collected or is missing). For subjects who are off study treatment, AEs will be included if event occurred within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment. No "subtracting rule" will be applied when an AE occurs both pre-treatment and post-treatment with the same preferred term and grade.
  - On-treatment evaluations (laboratory tests, pulse oximetry and vital signs) will be defined as evaluations taken after the day (and time, if collected and not missing) of first dose of study treatment. For subjects who are off study treatment, evaluations should be within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment.
- Late-emergent drug-related AEs will be defined as drug-related AEs with an onset date greater than 100 days after the last dose of study treatment in subjects who are off study treatment.

6.2 Treatment Regimens

The treatment group "**as randomized"** will be retrieved from the IRT system:

- Arm A: Experimental arm: nivolumab + cabozantinib
- Arm B: Experimental arm: nivolumab + ipilimumab + cabozantinib
- Arm C: Control arm: sunitinib

The treatment group "**as treated**" will be the same as the arm as randomized by IRT. However, if a subject received the incorrect drug for **the entire period** of treatment, the subject's treatment group will be defined as the incorrect drug the subject actually received.

Unless otherwise specified, the safety analysis will be based on the treatment group "as treated".

Unless otherwise specified, the efficacy analysis will be based on the treatment group "as randomized".

6.3 Populations for Analyses

All analyses will be performed using the treatment arm as randomized (intent to treat), with the exception of dosing and safety, for which the treatment arm as received will be used. For purposes of analysis, the following populations are defined in Table 6.3-1, and all populations for analyses given in this table refer to those subjects in Arm A and Arm C. Those subjects who randomized to Arm B prior to Revised Protocol 01 will be considered as part of the population of interest for descriptive summary of efficacy and safety analyses.

Table 6.3-1: Populations for Analyses

Population	Description
All Enrolled Subjects	All subjects who sign informed consent and were registered into the IRT.
All Randomized Subjects	All subjects who were randomized will be used for analyses of demography, protocol deviations, baseline characteristics, primary efficacy analysis, secondary efficacy analyses, and outcome research analysis which will be performed for this population.
All Treated Subjects	All subjects who received at least one dose of any study medication. This is the primary population for exposure and safety analyses.
Intermediate/Poor Risk Subjects	All subjects who were randomized with baseline IMDC prognostic score ≥1 at the time of randomization (per IRT). This population will be used for subset analyses of demography, protocol deviations, baseline characteristics, primary efficacy analysis, and secondary efficacy analyses on intermediate/poor risk subjects.
All Intermediate/Poor Risk Treated Subjects	All intermediate/poor risk subjects who received any dose of study therapy. This population will be used for subset analyses of exposure and safety analyses on intermediate/poor risk subjects.
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, discrete variables will be tabulated by the frequency and proportion of subjects falling into each category, grouped by treatment. Percentages given in these tables will be rounded to the first decimal and, therefore, may not always sum to 100%. Percentages less than 0.1 will be indicated as '< 0.1'. Continuous variables will be summarized by treatment group using the mean, standard deviation, median, minimum, and maximum values.

Time-to-event variables (e.g. time-to resolution) will be analyzed using the Kaplan-Meier technique. When specified, the median will be reported along with 95% CI using Brookmeyer and Crowley method[5] (using log-log transformation for constructing the confidence intervals[6]).

Unless otherwise specified, the stratified log-rank test will be performed to test the comparison between time to event distributions. Unless otherwise specified, the stratified hazard ratio between 2 groups along with CI will be obtained by fitting a stratified Cox model with the group variable as a unique covariate.

Confidence intervals for binomial proportions will be derived using the Clopper-Pearson method.[7] The unweighted difference in ORRs between the two treatment arms and corresponding asymptotic 95% CI will be estimated using a Newcombe method.[8]

P-values from sensitivity analyses for efficacy endpoints are for descriptive purpose only and there will be no multiplicity adjustment for these analyses.

The conventions to be used for imputing missing and partial dates for analyses requiring dates are described in Section 8.

Note that, the outputs will present Arm A, Arm B, and Arm C in the following way:

- All analysis specified in Section 7.3 and 7.4 will present A, B, and C together.
- For purposes of efficacy analyses, all randomized subjects in Arm A and Arm C will be tabulated and presented in the same table. A few seperate efficacy summary tables will be generated for those subjects randomized to Arm B, which will be specified in the corresponding sections.
- For purposes of safety analyses, all treated subjects in Arm A and Arm C will be tabulated and presented in the same table. A few separate safety summary tables will be generated for those subjects treated in Arm B, which will be specified in the corresponding sections.
- Summaries of outcome research and biomarker data will be tabulated on the Arm A and C subjects only.

7.1.1 Adverse Events, Serious Adverse Events, Multiple events, Select Adverse Events, Other Events of Special Interest and Immune-Mediated Adverse Events

Drug-related AEs are those events with relationship to study drug “Related”, as recorded on the CRF. If the relationship to study drug is missing, the AE will be considered as drug-related.

Serious adverse events consist of AEs deemed serious by the Investigator and flagged accordingly in the CRF and clinical database.

Adverse events leading to study drug discontinuation are AEs with action taken regarding study drug(s) = “Drug was discontinued”.

Adverse events leading to dose delay are AEs with action taken regarding study drug(s) = “Drug was delayed”.

Adverse event that led to dose delay of the oral drug (similarly defined as dose omission or dose interruption) will be coded with action “Drug was interrupted”.

Adverse events leading to dose reduction are AEs with action taken regarding study drug(s) = “Dose was reduced”.

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), and the most recent version of the dictionary at the time of the database lock will be used. Adverse events results will be graded for severity using NCI Common Terminology Criteria for Adverse

Events (CTCAE) and the most recent version of the criteria at the time of the database lock will be used.

In the AE summary tables, unless otherwise specified, subjects will be counted only once at the Preferred Term (PT), only once at the System Organ Class (SOC), and only once at subject level for the counting of total number of subjects with an AE. The AE tables will be sorted by the SOCs and then PTs. SOC will be ordered by descending frequency overall and then alphabetically. PTs will be ordered within SOC by descending frequency overall and then alphabetically. The sorting will be done based on the 'Any Grade' column of the experimental arm when arms are presented side-by-side.

Unless otherwise specified, the AE summary tables will be restricted to on-treatment events regardless of the causality.

Analyses that take into account the multiple occurrences of a given adverse event will be conducted (see Section 7.6.9). To prepare these analyses, the CRF data will be processed according to standard BMS algorithms[9] in order to collapse adverse event records into unique records based on the preferred term. These data will be presented as the rate per 100 person-years of exposure. These analyses will take into account all on-treatment events (allowing more than 1 event per subject) and the total exposure time. The person-year exposure will be computed as the sum over the subjects' exposure expressed in years where the exposure time is defined as

- (Date of last dose of study treatment - date of first dose of study treatment + 31 days (or 101 days, depending on the analysis))/365.25, for subject who are off study treatment and were followed for at least 30 days (or 100 days, depending on the analysis) after last dose of study treatment.
- (Last known alive date - date of first dose of study treatment +1)/365.25, for subjects who are still on-treatment or who are off study treatment and were followed less than 30 days (or 100 days depending on the analysis) after last dose of study treatment.

7.1.1.1 Select Adverse Events (EU/ROW Submissions)

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (e.g. pulmonary events, gastrointestinal events categories, etc.). AEs that may differ from or be more severe than AEs caused by non-immunotherapies and AEs whose early recognition and management may mitigate severe toxicity are included as select AEs. Categories of select AEs may include subcategories (e.g. adrenal disorders, diabetes, pituitary disorders, and thyroid disorders are subcategories of the endocrine event category).

The list of MedDRA preferred terms used to identify select adverse events is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock will be provided by categories/subcategories.

In addition to the frequency and worst severity of select AEs, time-to onset, time-to resolution, and time-to resolution where immune modulating medication was initiated will be analyzed for each specific category/subcategory of drug-related select AEs when applicable.

Further details on the definitions time-to onset and time-to resolution are described in APPENDIX 1.

7.1.1.2 Other Events of Special Interest

Other events of special interest (OEOSI) consist of a list of preferred terms grouped by specific category (e.g. Myositis Event, Myocarditis Event, Demyelination Event, Guillain-Barre Syndrome, Pancreatitis Event, Uveitis Event, Encephalitis Event, Myasthenic Syndrome, Rhabdomyolysis Event, Graft Versus Host Disease). The list of MedDRA preferred terms used to identify OEOSI is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock by categories will be provided.

7.1.1.3 Immune-Mediated Adverse Events (US Submission)

In order to further characterize AEs of special clinical interest, analysis of immune-mediated AEs (IMAE) will be conducted. IMAEs are specific events (or groups of PTs describing specific events) that include pneumonitis, diarrhea/colitis, hepatitis, nephritis/renal dysfunction, rash, endocrine (adrenal insufficiency, hypothyroidism/thyroiditis, hypothyroidism, thyroiditis, hyperthyroidism, diabetes mellitus, and hypophysitis), and other specific events, considered as potential immune-mediated events by investigator that meet the definition summarized below:

- those occurring within 100 days of the last dose,
- regardless of causality,
- treated with immune-modulating medication (of note, endocrine AEs such as adrenal insufficiency, hypothyroidism/thyroiditis, hypothyroidism, thyroiditis, hyperthyroidism, diabetes mellitus, and hypophysitis are considered IMAEs regardless of immune-modulating medication use, since endocrine drug reactions are often managed without immune-modulating medication).
- with no clear alternate etiology based on investigator assessment, or with an immune-mediated component

The list of MedDRA preferred terms used to identify IMAEs is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock by categories will be provided.

7.1.2 Laboratory Tests

Clinical laboratory parameters (hematology, serum chemistry and electrolytes) will be evaluated.

Laboratory tests will be graded using the NCI Common Terminology Criteria, and the most recent version of the criteria at the time of the database lock will be used.

Clinical laboratory data will be first analyzed using International System of Units (SI).

Analyses will be repeated using US conventional units.

In the laboratory summary tables, unless otherwise specified, subjects will be counted only once for each lab parameter according to their worst on treatment CTC grade (worst being the highest

CTC grade). The laboratory tables and listings will be sorted by laboratory category, laboratory subcategory and laboratory test code sequence number.

[redacted]

7.2 Study Conduct

The following programmable deviations will be considered as relevant protocol deviations and summarized by treatment group and overall in all randomized subjects. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and non-programmable) eligibility and on-treatment protocol deviations will be reported through ClinSIGHT listings.

Eligibility:

- Subjects with baseline KPS < 70%
- Subjects who received prior systemic anti-cancer treatment in the metastatic setting
- Subjects without histologically confirmed RCC with a clear-cell component, documented advanced or metastatic RCC

On-study:

- Subjects receiving anti-cancer therapy (chemotherapy, hormonal therapy, immunotherapy, standard or investigational agents for treatment of cancer) while on study therapy
- Subjects treated differently than as randomized (subjects who received the wrong treatment, excluding the never treated)

Enrollment by country and site, and enrollment by month will be summarized and listed for all enrolled subjects.

A by-subject listing of batch numbers for all treated subjects will be provided.

7.3 Study Population

Analyses in this section will be tabulated for all randomized subjects by treatment group as randomized, unless otherwise specified.

7.3.1 Subject Disposition

The total number of subjects enrolled (randomized or not randomized) will be presented along with the reason for not being randomized. This analysis will be performed on the all enrolled subjects population only.

Number of subjects randomized but not treated along with the reason for not being treated will be tabulated by treatment group as randomized.

Number of subjects who discontinued study treatment along with corresponding reason will be tabulated by treatment group as treated. Reason for discontinuation will be derived from subject status CRF page. This analysis will be performed only on the all treated subjects population.

A by-subject listing for all treated subjects will be provided showing the subject's off treatment date and whether the subject continue in the study along with the reason for going off study. A by-subject listing for all enrolled subjects will also be provided, showing whether the subject was randomized along with the reason for not being randomized.

Note that for all intermediate/poor risk treated subjects in Arm A and Arm C, number of subjects who discontinued study treatment along with corresponding reason will be tabulated by treatment group as treated.

7.3.2 Demographics and Other Baseline Disease Characteristics

The following demographic and baseline disease characteristics will be summarized and listed by treatment group as randomized:

- Age
- Age categorization (< 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85, ≥ 75, ≥ 65)
- Sex (Male, Female)
- Race
- Region (Region (US/Canada/W.Europe/N.Europe vs. ROW) (source: IRT)
- Ethnicity (Hispanic/Latino and Not Hispanic/Latino)
- Karnofsky performance status (70, 80, 90, 100)
- Baseline IMDC prognostic score (0, 1-2, ≥ 3) (source: IRT)
- Baseline IMDC prognostic score (0, 1-2, ≥ 3) (source: CRF)
- Time from initial disease diagnosis to randomization (<1 year, ≥1 year)
- Baseline LDH level (≤ 1.5 x ULN, >1.5 x ULN)
- Hemoglobin (<LLN, ≥ LLN)
- Corrected Calcium (≤ 10 mg/dl, >10mg/dl)
- Absolute Neutrophil Count (≤ ULN, > ULN)
- Platelet Count (≤ ULN, > ULN)
- Baseline Alkaline phosphatase (< ULN, ≥ ULN)
- Prior nephrectomy (Yes, No)
- Prior radiotherapy (Yes, No)
- Baseline PD-L1+ status based on a 1% cut off (≥ 1% vs. < 1% or indeterminate)
- Baseline PD-L1+ status based on a 5% cut off (≥ 5% vs. < 5% or indeterminate)
- Baseline PD-L1+ status based on a 10% cut off (≥ 10% vs. < 10% or indeterminate)

- Number of disease sites per subject (1, 2, 3, 4 , >4)
- Tumor burden: sum of the diameters of target lesions at baseline
- Most common sites of metastasis
- Sarcomatoid features (Yes, No)
- Stage at the initial diagnosis (Stage IV, non-Stage IV)
- Pre-treatment events tumor assessment (per Investigator)

Summary table (cross-tabulation) by treatment group for stratification factor (except for region) will be provided to show any discrepancies between what was reported through IRT vs. CRF at baseline. This summary will be performed based on all randomized subjects.

- IMDC Prognostic Score: 0 versus 1-2 versus 3-6
- Region: US/Canada/W Europe/N Europe versus ROW
- PD-L1 tumor expression: ≥ 1% versus < 1% or indeterminate

A listing of randomization scheme presenting randomized treatment group and as treated treatment group will be provided for all randomized subjects.

Note that for all intermediate/poor risk subjects in Arm A and Arm C, demographic and baseline disease characteristics will be summarized and listed by treatment group as randomized.

7.3.3 Medical History

A by-subject listing of general medical history for all randomized subjects will be provided.

7.3.4 Prior Therapy Agents

Prior adjuvant or neo-adjuvant therapy will be summarized by treatment group and overall.

Prior systemic cancer therapy will be summarized by treatment group and overall and listed by subject.

Prior radiotherapy and prior surgery related to cancer will be listed by subject.

7.3.5 Physical Examinations

Subjects with abnormal baseline physical examination will be listed by subject.

7.3.6 Baseline Physical Measurements

Baseline physical measurements will be listed by subject.

7.4 Extent of Exposure

Listings will include all available exposure data. Analyses will be performed by treatment group "as treated" in all treated subjects, unless otherwise specified.

7.4.1 Administration of Study Therapy

The following parameters will be summarized (descriptive statistics) by study therapy and treatment group:

- Number of doses received
- Cumulative dose
- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%
- Average daily dose

Duration of study therapy will be summarized (descriptive statistics) by treatment group.

A by-subject listing of dosing of study medication (record of study medication, infusion details, and dose changes) will be also provided.

Note that similar study therapy table will be summarized for all intermediate/poor risk treated subjects in Arm A and Arm C.

Table 7.4.1-1: Study Therapy Parameter Definitions for Arm A and C

	Nivolumab	Cabozantinib	Sunitinib
Dosing Schedule per Protocol	240 mg every 2 weeks	40 mg PO once daily	50 mg PO once daily for 4 weeks followed by 2 weeks off.
Dose	mg	mg	mg
Cumulative Dose	mg sum of the doses administered to a subject	mg sum of the doses administered to a subject	mg sum of the doses administered to a subject
Relative Dose Intensity (%)	[Cum dose (mg) /((Last dose date– First dose date + 14) × 240/14)] × 100	See below	See below
Duration of Treatment	*Last dose date - Start dose date +1*	*Last dose date - Start dose date + 1*	*Last dose date - Start dose date +15*

Additional Parameters -Cabozantinib treatment

Average daily dose (in mg/day) is defined as:

Sum of all Cabozantinib doses in mg actually received / duration of treatment in days.

Since Cabozantinib treatment consists of 40 mg PO daily dose, the planned dose intensity of Cabozantinib is 40 mg/day.

Relative dose intensity for Cabozantinib (%) is defined as: *(Average daily dose / 40) x 100.*

Additional Parameters -Sunitinib treatment

Average daily dose (in mg/day) is defined as:

Sum of all Sunitinib doses in mg actually received / duration of treatment in days

Since Sunitinib treatment consists of 50 mg PO daily dose for 4 weeks followed by 2 weeks of washout period, the planned dose intensity of Sunitinib is 33.33 mg/day (50 mg x 28 days / 42 days).

Relative dose intensity for Sunitinib (%) is defined as: *(Average daily dose / 33.33) x 100.*

Table 7.4.1-2: Study Therapy Parameter Definitions for Arm B (Cycle 1-4)

	Nivolumab	Ipilimumab	Cabozantinib
Dosing Schedule per Protocol	3 mg/kg every 3 weeks for 4 doses	1 mg/kg every 3 weeks for 4 doses	40 mg PO once daily
Dose	Dose (mg/kg) is defined as Total Dose administered (mg)/Most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.	Dose (mg/kg) is defined as Total Dose administered (mg)/Most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.	mg
Cumulative Dose	Cum Dose (mg/kg) is the sum of the doses administered to a subject.	Cum Dose (mg/kg) is the sum of the doses administered to a subject.	mg sum of the doses administered to a subject
Cycle Duration$_{(i)}$ (wk)	(Dose date$_{(i+1)}$ - Dose date$_{(i)}$)/7	(Dose date$_{(i+1)}$ - Dose date$_{(i)}$)/7	N/A
Cycle Intensity$_{(i)}$ (mg/kg/wk)	Dose$_{(i)}$/Cycle Duration$_{(i)}$	Dose$_{(i)}$/Cycle Duration$_{(i)}$	N/A
Relative Cycle Intensity $_{(i)}$ (%)	(Cycle Intensity$_{(i)}$/intended dose per week)(i) * 100	(Cycle Intensity$_{(i)}$/intended dose per week)(i) * 100	N/A
Relative Dose Intensity (%)	Sum of all Relative Cycle Intensities divided by N	Sum of all Relative Cycle Intensities divided by N	(Average daily dose / 40) x 100
Duration of Treatment	*Last dose date - Start dose date +1*	*Last dose date - Start dose date +1*	*Last dose date - Start dose date +1*

Table 7.4.1-3: Study Therapy Parameter Definitions for Arm B (Cycle 5 Onward)

	Nivolumab	Cabozantinib
Dosing Schedule per Protocol	240 mg every 2 weeks	40 mg PO once daily
Dose	mg	mg
Cumulative Dose	mg sum of the doses administered to a subject	mg sum of the doses administered to a subject
Relative Dose Intensity (%)	[Cum dose (mg) /((Last dose date– First dose date + 14) × 240/14)] × 100	(Average daily dose / 40) x 100
Duration of Treatment	*Last dose date - Start dose date +1*	*Last dose date - Start dose date + 1*

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delays

Each nivolumab infusion or sunitinib dose may be delayed. A dose will be considered as actually delayed if the delay is exceeding 3 days (ie greater than or equal to 4 days from scheduled dosing date) for nivolumab. Reason for dose delay will be retrieved from CRF dosing pages. It is worth noting that during the two week mandatory washout period for sunitinib, a daily dose of 0 mg will be entered in the CRF pages, with corresponding reason for dose modification recorded as "No Change".

If cabozantinib is given daily, a daily dose of 0 mg entered in the CRF pages will be considered as delay. If cabozantinib is given every other day, then a daily dose of 0 mg will be entered every other day in the CRF pages, with corresponding reason for dose modification recorded as "No Change". If there are more than one 0 mg daily dose entered consecutively, then this will be considered as delay.

The following parameters will be summarized by treatment group:

- Number of subjects with at least one dose delayed, the number of dose delays per subject, the reason for dose delay and the length of dose delay.

Note that similar dose delay summary table for Arm A and Arm C will be summarized for all intermediate/poor risk treated subjects.

For Arm B, both nivolumab and ipilimumab can be delayed at the same cycle. A dose will be considered as actually delayed if the delay is exceeding 3 days (ie greater than or equal to 4 days from scheduled dosing date) for nivolumab and ipimumab. Cabozantinib a daily dose of 0 mg entered in the CRF pages will be considered as delay. Similar table will be produced for Arm B

subjects previously randomized to Arm B continue with Arm B treatment and continue with Arm B clinical planned events, per protocol.

7.4.2.2 Infusion Interruptions and Rate Changes

Each nivolumab or ipilimumab infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment group:

- Number of subjects with at least one dose infusion interruption, the reason for interruption, and the number of infusion interruptions per subject.
- Number of subjects with at least one IV infusion rate reduction, the reason for reduction and the number of infusion with IV rate reduction per subject.

Note that similar summary table for Arm A and Arm C will be summarized for all intermediate/poor risk treated subjects.

7.4.2.3 Dose Escalations

Dose escalations are permitted for cabozantinib and sunitinib but not for nivolumab and ipilimumab.

7.4.2.4 Dose Reductions

Dose reductions are permitted for cabozantinib and sunitinib but not for nivolumab and ipilimumab.

Dose reduction for subjects treated with sunitinib is defined as at least one day with a non zero dose smaller than 50 mg and smaller than previous non zero dose with a CRF reason different from "Dosing Error" or "No Change".

Dose reduction for subjects treated with cabozantinib is defined as at least one day with 20 mg or 20 mg every other day with a CRF reason different from "Dosing Error" or "No Change".

Note that similar dose redcution summary table for Arm A and Arm C will be summarized for all intermediate/poor risk treated subjects.

The following summaries will be presented for the cabozantinib component of study treatment:

i) For dose reductions due to AE
   Categorical summaries for:

   - Subjects with any dose reduction
   - Dose levels received by a subject
   - Lowest non-zero dose level received
   - Last non-zero dose level received
   - Last dose level received (including dose holds)

Descriptive statistics for:

- Duration of treatment in months for each dose level (40 mg, 20 mg, 0 mg)
- Time to second dose level reduction (first receipt of 20mg) (days)

ii) Summaries for dose holds due to AE (those with 0 mg due to AE):

- Descriptive statistics for number of dose holds due to an AE
- Descriptive statistics for duration of dose holds per dose hold and per subject due to an AE, calculated as (stop date of hold – start date of hold + 1)
- Categorical summary for subjects with duration of holds due to an AE that can be classified as any number of days, ≥ 7 days, ≥ 14 days, ≥ 21 days, and >42 days
- Descriptive statistics for time to first dose hold, time to first dose hold that ≥ 7 days, ≥ 14 days, ≥ 21 days, and >42 days. The time to dose hold is calculated as (start date of the hold – first dose date + 1)
- Descriptive statistics for time to second dose hold, time to second dose hold that was ≥ 7 days, ≥ 14 days, ≥ 21 days, and >42 days.

iii) Summaries for dose modifications (defined as a reduction or hold) due to AE:

- Frequency counts and percentages for subjects with any dose modifications
- Descriptive statistics for number of dose modifications (0-3)
- Descriptive statistics for time to the first dose modification
- Descriptive statistics for time to the second dose modification

7.4.2.5 Dose Omissions

Dose omissions are not permitted.

[redacted]

7.5 Efficacy

Analyses in this section will be tabulated for all randomized subjects in Arm A and Arm C, unless otherwise specified. A few seperate efficacy summary tables will be generated for those subjects who randomized to Arm B prior to Revised Protocol 01, which is specified in the relevant subsections below.

Principal analyses of progression free survival (PFS) and objective response rate (ORR) will be based on the Blinded Independent Central Review (BICR) evaluation, unless noted otherwise.

Analyses in this section will be tabulated for all randomized subjects by treatment group as randomized, unless otherwise specified.

Unless stated otherwise, whenever a stratified analysis is specified, the following stratifications factors (recorded at randomization as per IRT) will be used:

- IMDC Prognostic Score: 0 versus 1-2 versus 3-6
- Region: US/Canada/W Europe/N Europe versus ROW
- PD-L1 tumor expression: ≥ 1% versus < 1% or indeterminate

The key secondary objective Overall Survival among all randomized subjects will be tested after conducting the primary objective analyses of PFS on all randomized subjects. For assessing this secondary objective of this study, a hierarchical testing procedure[10] will be used so that the overall experiment-wise Type I error rate is two-sided 0.05.

The secondary objective of ORR among all randomized subjects will be tested after conducting the key secondary objective analyses of OS on all randomized subjects, per a hierarchical testing procedure so that the overall experiment-wise Type I error rate is two-sided 0.05. Note that the formal ORR analysis specified in the SAP supersede those specified in the protocol.

Confidence intervals (CI) for primary and secondary endpoint analyses included in hierarchy will be based on nominal significance level adjusted for primary endpoints and interim analyses to preserve overall type one error rate.

Alpha (α) for the CI will be the same as nominal significance level for hypothesis testing. CIs for other endpoints will be at the two-sided 95% level. All p-values reported will be two-sided. P-values will be rounded to the fourth decimal place. Point estimates and confidence bounds for efficacy variables will be rounded to the second decimal place.

A by-subject listing of efficacy results will be presented including treatment group, treatment duration, BICR progression date, overall survival, death date, etc.

7.5.1 Analysis of Progression-Free Survival

The primary objective of the study is to compare the PFS per BICR of Arm A to Arm C in randomized subjects with previously untreated (first line) advanced or metastatic RCC. All the analyses outlined in this section are specified for the all randomized subjects population in Arm A and Arm C, unless otherwise specified.

PFS per BICR will be compared between the treatment groups via stratified log-rank test among all randomized subjects at a two-sided $\alpha = 0.05$ level. The stratification factors will be IMDC prognostic risk score (0 vs 1-2 vs 3-6), region (US/Canada/W Europe/N Europe vs ROW)' and PD-L1 status (≥ 1% vs < 1% or indeterminate).

The primary definition of PFS adjusting for subsequent anticancer therapy will be used in this analysis. The two-sided log-rank p-value will be reported.

The estimate of the PFS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio will also be presented.

The PFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood[11] formula for variance derivation and on log-log transformation applied on the survivor function[12].

Analyses of PFS will also be conducted based on the secondary definition of PFS. These analyses will be the same as those specified above.

The source of PFS event (progression or death) will be summarized by treatment group. The status of subjects who are censored (as per primary definition of PFS) in the PFS KM analysis will be tabulated for each treatment group including the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)
- No baseline tumor assessment
- No on-study tumor assessment and no death
- Received subsequent anticancer therapy

A by-subject listing will be presented including treatment group, PFS duration under the primary definition, PFS duration on the ITT definition, whether the subject was censored under the primary definition, and if censored, the reason, and whether the subject was censored under the ITT definition, and if censored, the reason.

A by-subject listing of lesion evaluations per BICR will be presented.

Note that similar tables (KM plot, median PFS with its 95% CI, PFS rates, source of PFS event, and status of subjects who are censored) along with by-subject listing will be summarized for all intermediate/poor risk subjects in Arm A and Arm C.

For those subjects who randomized to Arm B prior to Revised Protocol 01, the PFS function will be estimated using the KM product limit method and reported seperately for Arm B. A two-sided

95% CI for median PFS will be computed via the log-log transformation method. A by-subject listing will be presented including treatment group, PFS duration under the primary definition, PFS duration on the ITT definition, whether the subject was censored under the primary definition, and if censored, the reason, and whether the subject was censored under the ITT definition, and if censored, the reason.

7.5.2 Supportive Analyses of Progression-Free Survival

The following sensitivity analyses will be conducted using both the primary and the secondary definition of PFS in all randomized subjects:

1) Delayed effect of immunotherapy interventions may cause a late separation in the PFS KM curves and non-proportional hazards. PFS (as determined by BICR) will be compared between treatment groups via two-sided 0.05 stratified weighted log-rank test among subjects. The primary definition of PFS will be used in this analysis. The two-sided stratified weighted log-rank p-value will be reported using G (rho = 0, gamma = 1) weights, in the terminology of Fleming and Harrington[13].
   a) The Fleming Harrington test can be unstable, so it is possible, though uncommon, that thep-value for this trial will not be estimable.
   b) The estimate of the PFS hazard ratio in the period before and following 6 months will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. In this model, period is a binary variable indicating pre- vs. post-6 months. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio will also be presented.
2) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization, which, by definition, will be balanced across treatment groups, will still be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include:
   a) Age categorization (< 65 vs. ≥ 65)
   b) Gender (Male vs. Female)
   c) Race
   d) Region (US/Canada/W.Europe/N.Europe vs. ROW)
   e) IMDC score (0 vs 1-2 vs 3-6)
   f) Karnofsky performance status (100-90, <90)
   g) Prior Nephrectomy (Yes, No)
   h) LDH level (≤ 1.5 x ULN , > 1.5 x ULN)
   i) Baseline PD-L1+ status based on a 1% cut off
   j) Number of organ with metastasis (1 vs. ≥2)

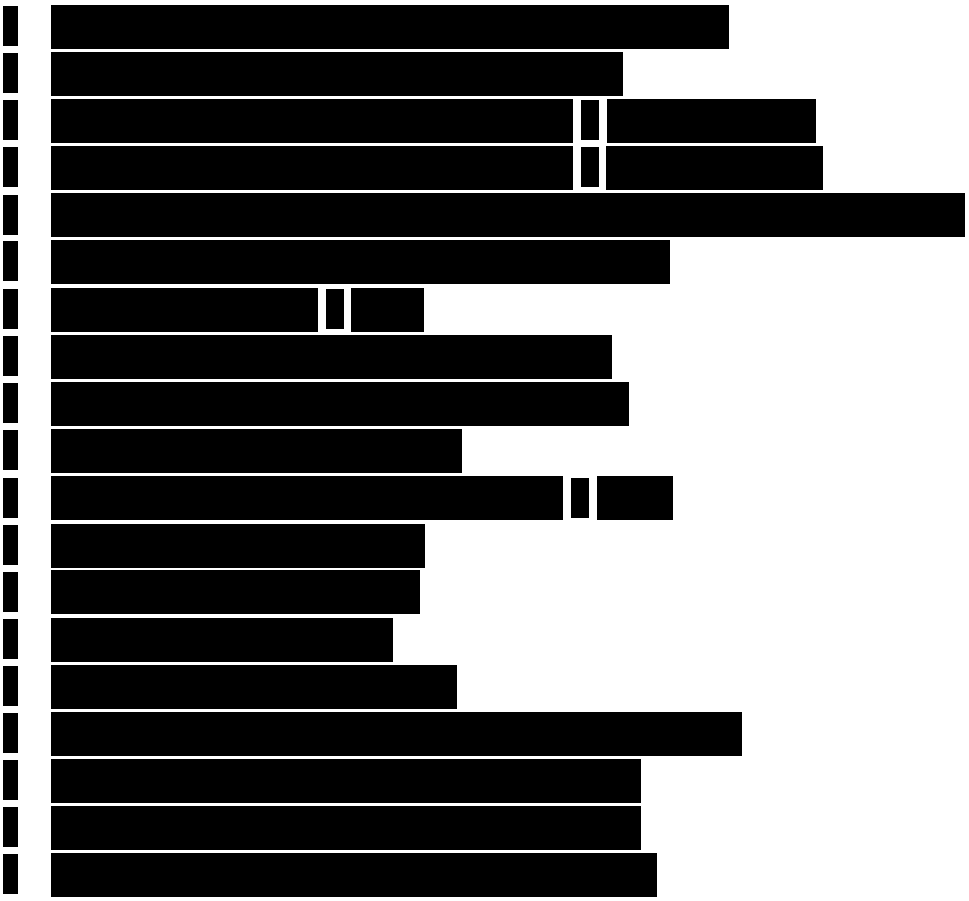
   The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 95% CIs.

3) PFS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.
4) PFS using the investigator's assessment. The hazard ratio associated with treatment and median PFS will be presented along with the associated two-sided 95% CIs.

   A cross tabulation of PFS assessment by BICR versus PFS assessment by investigator will be presented, by treatment group. Concordance Rate of event will be computed as the frequency with which investigator and BICR agree on classification of a subject as event versus censored as a proportion of the total number of randomized subjects assessed by both the investigator and BICR.

   A by-subject listing of PFS assessment per BICR and investigator will be presented.
5) PFS using an un-stratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.
6) PFS using an un-stratified Cox proportional hazards model, adjusted, using as covariates only the stratification factors used in randomization. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.
7) PFS for subjects with no relevant protocol deviations. This analysis will be conducted only if there are more than 10% subjects with relevant protocol deviations. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.
8) The method of Gail and Simon14 will be used to test for a qualitative interaction between treatment and strata. This test will be conducted at $\alpha = 0.10$ level. The p-value reported from this specific analysis is for descriptive purposes alone.
9) To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. A two-sided Wald Chi-square p-value of less than 0.1 may indicate a potential nonconstant treatment effect. [redacted].
10) PFS using censoring for 2 missing TA

[redacted]



7.5.4 Analysis of Overall Survival

One of the secondary objectives of the study is to compare the overall survival of Arm A to Arm C in all randomized subjects. All the analyses outlined in this section are specified for the all randomized subjects population in Arm A and Arm C, unless otherwise specified.

If the formal analysis of PFS among all randomized subjects is statistically significant, the formal interim analysis of OS among all randomized subjects will be tested, as per hierarchical testing procedure.

Overall survival will be compared between the treatment groups at the interim and final analyses, using stratified log-rank test. The stratification factors will be those used in the analysis of PFS. An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. The stratified hazard ratio between the treatment groups will be presented along with 100*(1- α)% CI (adjusted for interim). In addition, two-sided p-value will also be reported for the analysis of OS.

OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

The status of subjects who are censored in the OS KM analysis will be tabulated for each treatment group using the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)

A by-subject listing will be presented including treatment group, first and last dose date, whether the subject died, and if censored, the reason, event/censored date and OS duration.

Note that similar tables (KM plot, median OS with its 95% CI, OS rates, and status of subjects who are censored) along with by-subject listing will be summarized for all intermediate/poor risk subjects in Arm A and Arm C.

The analysis performed for PFS (detailed in section 7.5.8) for those subjects who randomized to Arm B prior to Revised Protocol 01 will be repeated for OS.

[redacted]

7.5.6 Current Status of PFS and OS Follow-up

The extent of follow-up for survival, defined as the time between randomization date and last known alive date (for subjects who are alive) or death date (for subjects who died), will be summarized descriptively (median, min, max, etc.) in months for all randomized subjects.

The currentness of follow-up for survival, defined as the time between last OS contact (i.e., last known alive date or death date) and cutoff date (defined by last subject last visit date), will be summarized in months for all randomized subjects. Subjects who died and subjects with last known alive date on or after data cut-off date will have zero value for currentness of follow-up.

Minimum follow-up of OS for all randomized subjects, defined as the time from cutoff date to last subject's randomization date, will be summarized in months.

Time from last evaluable tumor assessment to cutoff date in months will be summarized by treatment group and overall for all randomized subjects. Subjects who have a PFS event will be considered as current for this analysis. The secondary definition of PFS will be used for this summary.

In addition, time to treatment discontinuation will be summarized and presented by treatment group using a Kaplan-Meier curve whereby the last dose date will be the event date for those subjects who are off study therapy. Median duration of study therapy and associated 95% CI will be provided. Subjects who are still on study therapy will be censored on their last dose date.

A by-subject listing will also be produced to accompany the subject time from last evaluable tumor assessment.

7.5.7 Interim Analysis of Overall Survival

An independent statistician external to BMS will perform the analysis. In addition to the formal planned interim analyses for OS, the Data Monitoring Committee (DMC) will have access to periodic un-blinded interim reports of efficacy and safety to allow a risk/benefit assessment. Details are included in the DMC charter.

Two interim analyses of OS are planned for this study. The first interim analysis of OS is planned at the time of final PFS analysis and expected after observing 165 deaths (approximately 65% of the targeted OS events) have been observed among all randomized subjects in Arm A and Arm C based on above accrual rate and the exponential distribution in each arm. These formal comparisons of OS will allow for early stopping for superiority, and the boundaries for declaring superiority will be derived based on the actual number of deaths using Lan-DeMets spending function with O'Brien and Fleming type of boundary in EAST version 6. If the first interim analysis is performed exactly at 165 deaths, the boundary in terms of statistical significance for declaring superiority would be 0.011 (HR=0.673 with 16 months improvement in median OS for the Arm A versus Arm C comparison (33 versus 49 months)). The second interim analysis of OS is expected after observing 211 deaths (approximately 83% of the targeted OS events) have been observed among all randomized subjects based on above accrual rate and the exponential distribution in each arm. The boundary for declaring superiority in terms of statistical significance for the second interim analysis after 211 events would be 0.025 (HR=0.734 with 12 months improvement in median OS for the doublet versus sunitinib comparison (33 versus 45 months). The boundary for declaring superiority in terms of statistical significance for the final analysis after 254 events would be 0.041 (HR=0.774 with 9.6 months improvement in median OS for the Arm A versus Arm C comparison (33 versus 42.6 months).

Note that if the analysis of PFS final analysis and first interim analysis OS is trigger with 8-months minimum follow-up on all randomized subjects, minimum 283 PFS events and 149 OS events, then the details of the first interim analysis of OS will be as follows:

- If the first interim analysis is performed exactly at 149 deaths, the boundary in terms of statistical significance for declaring superiority would be 0.007 (HR=0.643 with 18.3 months improvement in median OS for the Arm A versus Arm C comparison (33 versus 51.3 months)).

The DMC will review the safety and efficacy data from the informal interim analyses, BMS will remain blinded to these interim results and and will determine if the study should continue with or without changes or if accrual should be stopped. Subject enrollment will continue while waiting for the DMC's decisions.

The chair of the DMC and the sponsor can call an unscheduled review of the safety data.

If the formal analysis of PFS among all randomized subjects is statistically significant, the formal interim analysis of OS among all randomized subjects will be tested, as per hierarchical testing procedure.

At the time of the formal interim analysis for superiority of OS, the DMC may recommend continuing or stopping the trial. If the trial continues beyond the formal interim analysis, BMS will remain blinded to these interim results and the nominal critical point for the final OS analysis will be determined using the recalculated information fraction at the time of the interim analysis, as described above. The final OS hazard ratio and corresponding confidence interval will be reported whereby the confidence interval will be adjusted accordingly (i.e. using the recalculated nominal α level at the final analysis).

If the trial is stopped for superiority of OS at the interim, the p-value from the interim stratified log-rank test will be considered the final primary analysis result.

7.5.8 Analysis of Objective Response Rate

One of the secondary objectives of the study is to evaluate the objective response rate in all randomized subjects in Arm A and Arm C. All the analyses outlined in this section are specified for the all randomized subjects population in Arm A and Arm C, unless otherwise specified. If the formal analysis of OS among all randomized subjects is statistically significant, the formal analysis of ORR among all randomized subjects will be tested, as per hierarchical testing procedure. Note that the formal ORR analysis specified in the SAP supersede those specified in the protocol.

The number and percentage of subjects in each category of BOR per BICR (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or unable to determine [UTD]) will be presented, by treatment group. Estimates of response rate, along with its exact two-sided 95% CI by Clopper and Pearson[15] will be presented, by treatment group.

Similar analyses will be repeated based on the investigator's assessment of ORR. A cross tabulation of BICR best response versus the investigator best response will be presented, by treatment group and by response categories. Concordance Rate of Responders will be computed as the frequency with which investigator and BICR agree on classification of a subject as responder vs. non responder/UTD as a proportion of the total number of randomized subjects assessed by both the investigator and BICR.

The following subject-level graphics will also be provided:

- For the responders only, time courses of the following events of interest will be graphically displayed: tumor response, progression, last dose received, and death.

- For response evaluable subjects (randomized subjects with baseline and at least one on-study tumor assessment),
  - A bar plot showing the best % reduction from baseline in sum of diameter of target lesions based on BICR assessment for each subject will be produced (excluding assessments after PD and assessments after start of subsequent anti-cancer therapy).
  - A plot of individual time course of tumor burden change per BICR assessment will be produced.

A by-subject listing of best overall response will be presented including treatment group, best overall response per BICR and dates of CR/PR/progression.

A by-subject listing of per time point tumor response per BICR will be presented.

Note that similar tables along with by-subject listing will be summarized for all intermediate/poor risk subjects in Arm A and Arm C.

For those subjects who randomized to Arm B prior to Revised Protocol 01, estimates of response rate, along with its exact two-sided 95% CI by Clopper-Pearson method, will be computed per BICR and investigator. DOR and TTR will also be evaluated per BICR.

[redacted]

7.5.10 Time to Tumor Response and Duration of Response

The analyses specified in this section will be conducted for all treatment arms. Duration of response (DOR) and time to response (TTR) will also be evaluated for subjects who achieved confirmed PR or CR. The DOR for each treatment group will be estimated using the Kaplan-Meier (KM) product limit method and will be displayed graphically. A table will be produced presenting number of events, number of subjects involved, medians, and 95% CIs for the medians. Median values of DOR, along with two-sided 95% CI in each treatment group will be computed based on a log-log transformation method.

The status of subjects who are censored in the DOR KM analysis will be tabulated for each treatment group including the following categories:

- Ongoing follow-up (current [last scan within adequate window vs cutoff date], not current)
- Off-study (lost to follow-up, withdraw consent, never treated)
- Received subsequent anticancer therapy.

TTR, which does not involve censoring, will be summarized by treatment group in all responders using descriptive statistics.

Cumulative Response Rates will be tabulated for Week 8, Month 4, 6, 8, and 12, and overall response rate will be provided.

A by-subject listing will be presented including treatment group, best response, time to response, duration of response, whether the subject was censored for duration of response, and, if so, the reason.

[redacted]

7.6 Safety

Analyses in this section will be tabulated for all treated subjects by treatment group as treated, unless otherwise specified.

Analyses in this section will be tabulated for all treated subjects in Arm A and Arm C, unless otherwise specified. Limited selection of the summary tables will be generated for those subjects who randomized to Arm B prior to Revised Protocol 01, which is specified in the relevant subsections below.

7.6.1 Deaths

Deaths will be summarized by treatment group:

- All deaths, reasons for death.
- Deaths within 30 days of last dose received, reasons for death.
- Deaths within 100 days of last dose received, reasons for death.

A by-subject listing of deaths will be provided for the all enrolled subjects population.

Note that for all intermediate/poor risk treated subjects in Arm A and Arm C, deaths will be summarized by treatment group.

Similar tables will be presented for those subjects treated in Arm B.

7.6.2 Serious Adverse Events

Serious adverse events will be summarized by treatment group:

- Overall summary of SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

All analyses will be conducted using the 30-day safety window.

A by-subject SAE listing will be provided for the "enrolled subjects" population.

Note that for all intermediate/poor risk treated subjects in Arm A and Arm C, serious adverse events will be summarized by treatment group.

Similar tables will be presented for those subjects treated in Arm B.

7.6.3 Adverse Events Leading to Discontinuation of Study Therapy

AEs leading to discontinuation will be summarized by treatment group:

- Overall summary of AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

The analyses will be conducted using the 30-day safety window.

A by-subject AEs leading to discontinuation listing will be provided.

Note that for all intermediate/poor risk treated subjects in Arm A and Arm C, AEs leading to discontinuation will be summarized by treatment group.

Similar table will be presented for those subjects treated in Arm B.

Note that

- for Arm A, AEs leading to discontinuation from cabozantinib only, nivolumab only, and both cabozantinib and nivolumab.
- for Arm B, AEs leading to discontinuation from cabozantinib only, nivolumab and ipilimumab only, and from cabozantinib, nivolumab, and ipilumumab will be summarized in all the tables specified in this section.

7.6.4 Adverse Events Leading to Dose Modification

AEs leading to dose delay/reduction will be summarized by treatment group:

- Overall summary of AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

- Overall summary of related AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

The analysis will be conducted using the 30-day safety window.

A by-subject AEs leading to dose delay/reduction listing will be provided.

7.6.5 Adverse Events

Adverse events will be summarized by treatment group.

The following analyses will be conducted using the 30 days safety window only:

- Overall summary of any AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.
- Overall summary of any AEs presented by worst CTC grade (any grade, grade 3-4, grade 5) by SOC/PT. This table will be restricted to events with an incidence greater or equal to 5% in any treatment group.
- Overall summary of any non-serious AEs presented by SOC/PT. This table will be restricted to events with an incidence greater or equal to 5% in any treatment group.
- Overall summary of any AEs that required immune modulating medication by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.

The following analyses will be conducted using the 30 days safety window and repeated using the 100 days safety window:

- Overall summary of drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

A by-subject AE listing will be provided. A by-subject listing of any AE requiring immune modulating medications will also be provided.

For those subjects treated in Arm B and for those intermediate/poor risk treated subjects in Arm A and Arm C:

The following analyses will be conducted using the 30 days safety window only:

- Overall summary of any AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.
- Overall summary of any AEs that required immune modulating medication by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.

The following analyses will be conducted using the 30 days safety window and repeated using the 100 days safety window:

- Overall summary of drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

A by-subject AE listing will be provided. A by-subject listing of any AE requiring immune modulating medications will also be provided

7.6.6 Select Adverse Events (EU/ROW Submissions)

Unless otherwise specified, analyses will be performed by select AE category. Analyses will also be repeated by subcategory of endocrine events.

7.6.6.1 Incidence of Select AE

Select AEs will be summarized by treatment group for each category/subcategory.

The following analyses will be conducted using the 30-day safety window only:

- Overall summaries of any select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory/PT.
- Overall summaries of any drug-related select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory/PT.
- Overall summaries of any serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Overall summaries of drug-related serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Overall summaries of any select AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Overall summaries of drug-related select AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Summary of frequency of unique select AEs by Category.

A by-subject select AE listing will be provided.

7.6.6.2 Time-to Onset of Select AE

Time-to onset of drug-related select AEs (any grade, grade 3-5) will be summarized for each category/subcategory by treatment group.

Time-to onset analyses are restricted to treated subjects who experienced at least one drug-related select AE in the category/subcategory. The analyses will be conducted using the 30-day safety window.

Additional details regarding the time-to onset definition are described in time-to onset definition subsection of APPENDIX 1.

7.6.6.3 Time-to Resolution of Select AE

Time-to resolution of the following specific events will be summarized separately for each category/subcategory.

- Time-to resolution of drug-related select AE (any grade, grade 3-5) by treatment group
- Time-to resolution of drug-related select AE (any grade, grade 3-5) where immune modulating medication was initiated, by treatment group

Time-to resolution analyses are restricted to treated subjects who experienced the specific events. Time-to resolution where immune modulating medication was initiated analyses are restricted to treated subjects who experienced the specific events and who received immune modulating medication during the longest select AE.

The analyses will be conducted using the 30-day safety window.

The following summary statistics will be reported: percentage of subjects with resolution of the longest select AE, median time-to resolution along with 95% CI (derived from Kaplan-Meier estimation) and ranges.

See time-to resolution definition subsection of APPENDIX 1 for additional details.

7.6.7 Immune-Mediated Adverse Events (US Submission)

IMAEs will be summarized by treatment group for each immune-mediated category / PT using the 100-day safety window:

- Overall summary of non-endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT.
- Overall summary of endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Overall summary of non-endocrine IMAEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT.
- Overall summary of endocrine IMAEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Overall summary of non-endocrine IMAEs leading to dose delay or reduction by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT
- Overall summary of endocrine IMAEs leading to dose delay or reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Summaries of time-to onset and time-to resolution of non-endocrine IMAEs where immune modulating medication was initiated presented by Category.
- Summaries of time-to onset and time-to resolution of endocrine IMAEs presented by Category.

A by-subject listing of IMAEs will be provided. By-subject listings of time-to resolution for longest IMAEs cluster (any grade and grade 3-5 in separate summaries) will also be provided. For new studies which collect investigator assessment of potential IMAE data, a by-subject listing of AEs considered as immune-mediated events per investigator but not qualified for IMAEs definition will also be provided.

In addition, for all nivolumab treated subjects who experienced at least one IMAE, the following data presentation will be provided:

- Summary of subjects who were re-challenged with nivolumab by IMAE category, with extended follow-up
- Summary of subjects who were re-challenged with nivolumab or ipilimumab by IMAE category, with extended follow-up

For these, re-challenge is considered to have occurred when last nivolumab and/or ipilimumab infusion was administered after the onset of an IMAE.

7.6.8 Other Events of Special Interest

OEOSI will be summarized by treatment group for each category.

The following analyses will be conducted using the 100-day safety window:

- Overall summary of OEOSI by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT
- Overall summary of drug-related OEOSI by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT

A by-subject listing of OEOSI will be provided.

7.6.9 Multiple Events

The following summary tables will be provided:

- A table showing the total number and rate (exposure adjusted) of occurrences for all AEs.
- A table showing the total number and rate (exposure adjusted) of occurrences for AEs occurring in at least 5% of subjects in any treatment group.

In addition, the rate (exposure adjusted) and its 95% CI evaluated for different time intervals will be displayed graphically for each treatment group. This analysis will be limited to the rate of all AEs and all drug-related AEs. The analyses will be conducted using the 30-day safety window.

A listing displaying the unique instances of all AEs, i.e., after duplicates have been eliminated and overlapping and contiguous occurrences of the same event (i.e. same PT) have been collapsed will be provided. No formal comparisons will be made between treatment groups.

7.6.10 Laboratory Parameters

The analysis population for each laboratory test is restricted to treated subjects who underwent that laboratory test. Laboratory tests (in addition to the tests specified below) with CTC criteria collected in the specific studies may also be included in the summaries.

A by-subject listing of differences in categorization of SI and US laboratory test results will be provided.

7.6.10.1 Hematology

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: hemoglobin (HB), platelets, white blood counts (WBC), absolute neutrophils count (ANC) and lymphocyte count (LYMPH).

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these laboratory parameters will be provided.

7.6.10.2 Serum Chemistry

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: ALT, AST, alkaline phosphatase (ALP), total bilirubin and creatinine.

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these laboratory parameters will be provided.

7.6.10.3 Electrolytes

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: sodium (high and low), potassium (high and low), calcium (high and low), magnesium (high and low), and Glucose Serum (fasting hyperglycemia and hypoglycemia regardless of fasting status).

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these laboratory parameters will be provided.

7.6.10.4 Additional Analyses

In addition, further analyses on specific laboratory parameters will be performed by treatment group:

<u>Abnormal Hepatic Function Test</u>

The number of subjects with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- ALT or AST > 3 x ULN, > 5 x ULN, > 10 x ULN and > 20 x ULN
- Total bilirubin > 2 x ULN

- ALP > 1.5 x ULN
- Concurrent (within 1 day) ALT or AST > 3 x ULN and total bilirubin > 1.5 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 1.5 x ULN
- Concurrent (within 1 day) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these specific abnormalities will be provided.

Abnormal Thyroid Function Test

The number of subjects with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- TSH value > ULN and
  - with baseline TSH value ≤ ULN
  - with at least one FT3/FT4 test value < LLN within 2-week window after the abnormal TSH test
  - with all FT3/FT4 test values ≥ LLN within 2-week window after the abnormal TSH test
  - with FT3/FT4 missing within 2-week window after the abnormal TSH test.
- TSH < LLN and
  - with baseline TSH value ≥ LLN
  - with at least one FT3/FT4 test value > ULN within 2-week window after the abnormal TSH test
  - with all FT3/FT4 test values ≤ ULN within 2-week window after the abnormal TSH test
  - with FT3/FT4 missing within 2-week window after the abnormal TSH test

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these specific abnormalities will be provided.

7.6.11 Vital Signs and Pulse Oximetry

Vital signs and pulse oximetry (i.e. % oxygen saturation) collected on the CRF will be provided in separate listings.

7.6.12 Physical Measurements

Physical measurements will be listed by subject.

7.6.13 Non-Protocol Medical Procedures

Non-protocol medical procedures will be listed by subject.

[redacted]

7.6.15 Pregnancy

A by-subject listing of pregnancy tests results will be provided for randomized female subjects.

7.6.16 Adverse Events By Subgroup

Overall summary of any AEs and drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT and for each treatment group for the following subgroups:

- Sex (Male vs. Female)
- Race
- Age (< 65 vs. 65 - < 75 vs. 75 - < 85 vs. ≥ 85 vs. ≥ 75 vs. ≥ 65)
- Region (US/Canada/W.Europe/N.Europe vs. ROW)

These analyses will be conducted using the 30-day safety window only.

[redacted]

[redacted]

8 CONVENTIONS

The following conventions may be used for imputing partial dates for analyses requiring dates:

- For missing and partial adverse event onset dates, imputation will be performed using the Adverse Event Domain Requirements Specification[16]
- For missing and partial adverse event resolution dates, imputation will be performed as follows (these conventions may change):
  - If only the day of the month is missing, the last day of the month will be used to replace the missing day. If the imputed date is after the death date or the last known alive date, then the latest known alive date or death date is considered as the resolution date.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.

- Missing and partial non-study medication domain dates will be imputed using the derivation algorithm described in 4.1.3 of BMS Non-Study Medication Domain Requirements Specification[17].
- Missing and partial radiotherapy and surgery dates will be imputed using algorithm described in APPENDIX 2.
- For death dates, the following conventions will be used for imputing partial dates:
  - If only the day of the month is missing, the 1st of the month will be used to replace the missing day. The imputed date will be compared to the last known alive date and the maximum will be considered as the death date.
  - If the month or the year is missing, the death date will be imputed as the last known alive date.
  - If the date is completely missing but the reason for death is present, the death date will be imputed as the last known date alive.
- For date of progression after start of study therapy, the following conventions will be used for imputing partial dates:
  - If only the day of the month is missing, the 1st of the month will be used to replace the missing day. In case of the date of death is present and complete, the imputed progression date will be compared to the date of death. The minimum of the imputed progression date and date of death will be considered as the date of progression.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- For date of progression to prior therapies, the following conventions will be used for imputing partial dates:
  - If only the day of the month is missing, the 1st of the month will be used to replace the missing day.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- For other partial/missing dates, the following conventions were used:
  - If only the day of the month is missing, the 15th of the month will be used to replace the missing day.
  - If both the day and the month are missing, “July 1” will be used to replace the missing information.
  - If a date is completely missing, it will be considered as missing.

The following conversion factors will be used to convert days to months or years:

1 month = 30.4375 days and 1 year = 365.25 days.

Duration (e.g. time-to onset, time-to resolution) will be calculated as follows:

Duration = (Last date - first date + 1)

Last known alive date will be defined based on all appropriate dates collected on the CRF.

All statistical analyses will be carried out using SAS (Statistical Analysis System software, SAS Institute, North Carolina, USA) unless otherwise noted.

9 CONTENT OF REPORTS

All analyses described in this SAP will be included in the Clinical Study Report(s) except where otherwise noted. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

10 DOCUMENT HISTORY

Table 10-1: Document History

Version Number	[redacted]	Description
1	[redacted]	Original issue
2	[redacted]	Employ hierarchical strategy to ORR after PFS and OS. [redacted]. Wording of cabozantinib dose delay and dose reductions are cleaned.

APPENDIX 1 TIME-TO ONSET AND TIME-TO RESOLUTION DEFINITION AND CONVENTIONS FOR SELECT ADVERSE EVENTS, IMMUNE-MEDIATED ADVERSE EVENTS AND EVENTS OF SPECIAL INTEREST

Time-to onset definition

Time-to onset of AE (any grade) for a specific category is defined as the time between the day of the first dose of study treatment and the onset date of the earliest AE (of any grade) in this category.

The time-to onset of AE (grade 3-5) for a specific category is defined similarly with an onset date corresponding to a grade 3-5 AE.

Time-to onset of drug-related AE (any grade or grade 3-5) for a specific category is defined similarly but restricted to drug-related AE.

Time-to onset for a specific subcategory is defined similarly but restricted to event of this subcategory.

Time-to resolution definition

In order to derive the time-to resolution, overlapping or contiguous AEs within a specific category or subcategory will be collapsed into what will be termed "clustered" AEs. For example, if a subject (without pre-treatment AE) experienced an AE from 1st to 5th January, another AE (with different PT but within same category) from 6th to 11th January and same AE from 10th to 12th January, these will be collapsed into one clustered AE from 1st to 12th January.

Key derivation steps for each type of clustered AEs is summarized as follows:

- For any grade AE: Collapse any on-treatment AE from the same category

- For drug-related any grade AE: Collapse any on-treatment drug-related AE from the same category
- For grade 3-5 AE: Collapse any on-treatment AE from the same category. Resolution will be based on the onset date of the earliest grade 3-5 records (if no grade 3-5 record, clustered AE is excluded)
- For drug-related grade 3-5 AE: Collapse any on-treatment drug-related AE from the same category. Resolution will be based on the onset date of the earliest grade 3-5 records (if no grade 3-5 record, clustered AE is excluded)

Time-to resolution of AE (any grade) for a specific category is defined as the longest time from onset to complete resolution or improvement to the grade at baseline among all clustered AEs experienced by the subject in this category per adverse event criteria category. Events which worsened into grade 5 events (death) or have a resolution date equal to the date of death are considered unresolved. If a clustered AE is considered as unresolved, the resolution date will be censored to the last known alive date. Improvement to the grade at baseline implies that all different events in the clustered adverse event should at least have improved to the corresponding (i.e. with same preferred term) baseline grade. This measure is defined only for subjects who experienced at least one AE in the specific category.

The time-to resolution of AE (grade 3-5) for a specific category is defined similarly with an onset date corresponding to a grade 3-5 AE.

Time-to resolution of drug-related AE (any grade or grade 3-5) for a specific category is defined similarly but restricted to drug-related AE.

The time-to resolution of AE (any grade or grade 3-5, drug-related or all) where immune modulating medication was initiated is defined similarly. For data presentation not restricted to IMAE, the additional condition that the subject started an immune modulating medication during the longest AE resolution period will be applied.

Time-to resolution for a specific subcategory is defined similarly but restricted to event of this subcategory.

The algorithm for collapsing adverse event records is using the following conventions:

For each subject and specified category, the corresponding adverse event records will be collapsed when:

1) Multiple adverse event records have the same onset date.
2) The onset date of an event record is either the same day or 1 day later than the resolution date of a preceding event record (contiguous events).
3) The onset date of an event record is after the onset date and prior to or on the resolution date of a preceding event record (overlapping events).