STATISTICAL ANALYSIS PLAN

Study Protocol Number:	E7080-G000-307/KEYNOTE-581
Study Protocol Title:	A Multicenter, Open-label, Randomized, Phase 3 Trial to **C**ompare the Efficacy and Safety of **L**envatinib in Combination with **E**verolimus or Pembrolizumab Versus Sunitinib Alone in First-Line Treatment of Subjects with **A**dvanced **R**enal Cell Carcinoma (CLEAR)
Date:	14 Aug 2020
Version:	3.0

DATE	Highlights of Major Changes Sections/Changes
14 Aug 2020	**Version 3.0:**
	Section 3.1.3: Per protocol Amendment 07, delete "To assess PFS using immune-related RECIST (irRECIST) in subjects treated with lenvatinib in combination with pembrolizumab."
	Section 5.1.3: Per protocol Amendment 07, delete " Progression-free survival (PFS) as determined by IIR using irRECIST for subjects receiving lenvatinib plus pembrolizumab (Arm B), defined as the time from the date of randomization to the date of the first documentation of confirmed immune-related progressive disease (irPD) or death (whichever occurs first)."
	Section 5.2.1: Delete "or serum concentrations of pembrolizumab" in the definition of "Population Pharmacokinetic (PK) Analysis Set" for the correction.
	Section 5.2.3: "Major" was added for the correction. Only major protocol deviations will be listed.
	Section 5.3: Per protocol Amendment 07, delete "The tumor response based on irRECIST for Arm B only will be evaluated by IIR only after database lock for the planned final analysis for PFS."
	Section 5.4.3: Per protocol Amendment 07, delete the whole section of "Analysis pf Progression Free Survival Using irRECIST". "An analysis of overall concordance on PFS for each treatment group between assessment by IIR and by investigator review will be performed." was added.
	Section 5.7: "Protocol deviations and adverse events associated with COVID-19 will be presented" was added in section of "Other Analyses".
	Section 8.1.1: Per protocol Amendment 07, delete "Per irRECIST, irCR, irPR and irPD in arm B in this study require confirmation by a subsequent assessment of response at least 4 weeks (28 days) later. The irRECIST analysis for Arm B will be performed by IIR following the data base lock for the primary analysis."

DATE	Highlights of Major Changes Sections/Changes
15 Nov 2019	**Version 2.0:**
	Section 3.2: Target PFS events was updated according to Protocol Amendment 06 (PA06)
	Sections 4, 5.4.1, 5.4.2: Added an interim analysis of PFS projected to occur ~38 months after the first subject was randomized in study. The final PFS analysis is projected to occur approximately ~45 months after the first subject was randomized in the study. Based on the results from the two recent available IO+VEGF studies, an interim analysis of PFS is added. The number of interim analyses of OS is also increased from 1 to 3. The alpha-spending functions for PFS and OS analyses are added or updated. The procedure to control familywise error rate is changed to a graphical approach. The powers of statistical testing are re-calculated under the graphical approach. Updated median OS assumption for sunitinib arm from 30 months to 37.9 months and updated the projected timing for interim analyses and final analysis of OS accordingly. The projected timing for final analysis of OS is updated from 53 months to 69 months after first subject randomized.
	Section 5.1.2, 5.1.3 and 5.4.2: Updated to confirmed and unconfirmed PR/CR. The analysis of the confirmed CR/PR will be the primary analysis for ORR.
	Sections 5.2.1: Added Pembrolizumab PK Analysis Set
	Section 5.2.5: Updated the definition of prior medications to “Prior medications are defined as medications that started prior to the first dose of study drug regardless when the medications ended..”
	Section 5.3.3: Multiplicity adjustment strategy changed from truncated Hochberg procedure to Graphical approach with initial alpha of 0.045 assigned to test PFS lenvatinib+pembrolizumab vs sunitinib and initial alpha of 0.0049 assigned to test PFS lenvatinib+everolimus vs sunitinib. The multiplicity test strategy was updated to optimize the probability of success of all

DATE	Highlights of Major Changes Sections/Changes
	hypothesis tests while to strongly control the familywise type I error rate under 0.05 (2-sided).
	Section 5.3.4: Updated subgroups: KPS score group (100-90, 80-70) and number of metastatic sites per IIR (0, 1, 2, ≥3)
	Section 5.4.1: Removed the sensitivity analysis: "Using discontinuation of treatment due to reasons other than complete response and lost to follow up, or initiation of new anti-cancer treatment (whichever occurs later) to be a PD event for subjects without documented PD or death." Removed the sensitivity analysis: "A sensitivity analysis using a weighted log-rank test with Fleming-Harrington's $G^{\rho,\gamma}$ class of weights with $\rho = 0$ and $\gamma > 0$ will also be conducted for comparing PFS of lenvatinib + pembrolizumab (Arm B) versus sunitinib alone (Arm C) to take into consideration of possible delayed onset of treatment effect in lenvatinib + pembrolizumab arm." Added a sensitivity analysis: "Using the different derivation rule for the situation with more than one missed visit/tumor assessment: death or progression immediately after more than one missed visit/tumor assessment (i.e. if a subject missed two or more tumor assessments right before PD or death), the subject will be censored on the date of the last adequate tumor assessment before PD or death. Note that if a subject is censored by both this criterion and the anticancer treatment criterion, the earliest censoring date will be used."
	Section 5.4.2: Deleted "For subjects with no disease progression documented on second line therapy but have started third line therapy, the start date of third line therapy minus 1 day will be taken as PFS2 event date."
	Section 5.6.1.2: To align with summary conventions for pembrolizmab, the dose intensity, relative dose intensity for pembrolizumab was removed.

DATE	Highlights of Major Changes Sections/Changes
	Section 5.6.2, 5.6.2.2 Updated: "The serious AE that emerged up to 120 days after the subject's last dose of study drug is counted as TEAE". Added "SAE listing". Removed "cycle of worst grade" and "time to resolution". Removed two redundant examples of "Hypertension" and "Proteinuria", since these two baskets were described in general CSAE for lenvatinib summaries.
	Section 6: Updated Table 6 "Summary of Interim and Final Efficacy Analyses", to specify timing of interim and final analyses.

1 TABLE OF CONTENTS

2 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AEs	adverse events
AEOSI	adverse events of special interest
ATC	anatomical therapeutic chemical
BLQ	below limit of quantification
BMI	body mass index
BOR	best overall response
CBR	clinical benefit rate
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
COVID-19	coronavirus disease 2019
CR	complete response
CRF	case report form
CSAE	clinically significant adverse events
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DBP	diastolic blood pressure
DCR	disease control rate
DMC	data monitoring committee
DOR	duration of response
ECG	electrocardiogram
EORTC	European Organization for the Research and Treatment of Cancer
EuroQol	European quality of life
FDA	Food and Drug Administration
FKSI-DRS	functional assessment of cancer therapy kidney syndrome index-disease-related symptoms
FWER	family-wise error rate
HRQoL	health-related quality of life
IIR	independent imaging review
IMDC	International Metastatic Renal Cell Carcinoma Database Consortium
ITT	intent-to-treat data set
IV	Intravenous
K-M	Kaplan-Meier
KPS	Karnofsky performance status
LVEF	left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
MSKCC	Memorial Sloan-Kettering Cancer Center
NA	not applicable

Abbreviation	Term
NE	not evaluable
NYHA	New York Heart Association
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
Q3W	every 3 weeks
PK	pharmacokinetic
PR	partial response
PT	preferred term
RCC	renal cell carcinoma
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SI	Système International
SOC	system organ class
TEAEs	treatment-emergent adverse events
TEMAV	treatment-emergent markedly abnormal laboratory values
TLGs	tables, listings, and graphs
TNM	tumor, node, metastasis
ULN	upper limit of normal
WHO	World Health Organization

3 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report results for randomization phase of Eisai Protocol E7080-G000-307 or Merck KEYNOTE-581 Amendment 07 (Dated 06 Aug 2020) (hereafter referred to as Study 307). The SAP will be finalized prior to the database lock for the clinical study report.

The analysis plan for interim analysis of objective response rate (ORR) was prepared and finalized in a separate SAP. Additional plans for analyses of pharmacokinetics/pharmacodynamics (PK/PD), biomarkers, health-related quality of life (HRQoL), and relationships between PK/PD and efficacy/safety will be provided as separate SAPs.

3.1 Study Objectives

3.1.1 Primary Objective

The primary objective of the study is to demonstrate that lenvatinib in combination with everolimus (Arm A) or pembrolizumab (Arm B) is superior compared to sunitinib alone (Arm C) in improving progression-free survival (PFS) by independent imaging review (IIR) using Response Evaluation Criteria in Solid Tumors (RECIST 1.1) as first-line treatment in subjects with advanced renal cell carcinoma (RCC).

3.1.2 Secondary Objectives

The secondary objectives of this study are:

- To compare ORR by IIR using RECIST 1.1 of subjects treated with lenvatinib in combination with everolimus or pembrolizumab versus sunitinib.
- To compare overall survival (OS) of subjects treated with lenvatinib in combination with everolimus or pembrolizumab versus sunitinib.
- To compare safety and tolerability of treatment with lenvatinib in combination with everolimus or pembrolizumab versus sunitinib, including the assessment of the proportion of subjects who discontinued treatment due to toxicity and time to treatment failure due to toxicity.
- To compare the impact of treatment on HRQoL as assessed by using the Functional Assessment of Cancer Therapy Kidney Syndrome Index-Disease-Related Symptoms (FKSI-DRS), the European Organization for the Research and Treatment of Cancer (EORTC) QLQ-30 and the EuroQol EQ-5D-3L instruments, for subjects treated with lenvatinib in combination with everolimus or pembrolizumab versus sunitinib.
- To assess PFS on next-line of therapy (PFS2) as reported by investigator.
- To assess PFS based on investigator assessment per RECIST 1.1.
- To characterize the population pharmacokinetics (PK) of lenvatinib when co-administered with everolimus or pembrolizumab.

- To compare the PK of pembrolizumab from this study to historical data.
- To characterize the population PK of everolimus when co-administered with lenvatinib.
- To assess the PK/pharmacodynamic relationship between exposure and efficacy/ biomarkers/ safety, if possible, using a mechanistic approach.

3.1.3 Exploratory Objectives

The exploratory objectives of this study are:

- To compare ORR by investigator assessment using RECIST 1.1.
- To assess the duration of response (DOR) by IIR and investigator assessment using RECIST 1.1 for subjects in all treatment arms.
- To compare the disease control rate (DCR) (complete response [CR], partial response [PR] or stable disease [SD]) and clinical benefit rate (CBR) (CR, PR or durable SD) by IIR and investigator assessment using RECIST 1.1 of subjects treated with lenvatinib in combination with everolimus or pembrolizumab versus sunitinib.
- To compare PFS by IIR and investigator assessment using RECIST 1.1 in subjects treated with lenvatinib in combination with everolimus (Arm A) versus lenvatinib in combination with pembrolizumab (Arm B).
- To investigate the relationship between candidate tumor and blood biomarkers and clinical outcome measures including antitumor activity of study treatment.

3.2 Overall Study Design and Plan

This is a multicenter, randomized, open-label, Phase 3 study to compare the efficacy and safety of lenvatinib in combination with everolimus or pembrolizumab versus sunitinib as first-line treatment in subjects with advanced RCC.

Following the two stratification factors, i.e., geographic region (Region 1: Western Europe and North America or Region 2: rest of the world) and Memorial Sloan-Kettering Cancer Center (MSKCC) prognostic groups (favorable, intermediate and poor risk), Approximately 1050 eligible subjects will be randomized to one of the following three treatment arms in a 1:1:1 ratio, with approximately 350 subjects in each arm:

- Arm A: lenvatinib 18 mg (orally, once daily) plus everolimus 5 mg (orally, once daily)
- Arm B: lenvatinib 20 mg (orally, once daily) plus pembrolizumab 200 mg (intravenously [IV], every 3 weeks [Q3W])
- Arm C: sunitinib 50 mg (orally, once daily) on a schedule of 4 weeks on treatment followed by 2 weeks off (Schedule 4/2)

For the stratification factor of the geographic region, the Western Europe and North America region includes Austria, Belgium, Canada, France, Germany, Greece, Ireland, Italy, Netherlands, Poland, Spain, Switzerland, United Kingdom and the United States of America. The Rest of World region includes Australia, Czech Republic, Israel, Japan, Korea and Russian Federation.

This study is conducted in three study phases shown in Figure 1, a Pre-randomization Phase, a Randomization Phase, and an Extension Phase.

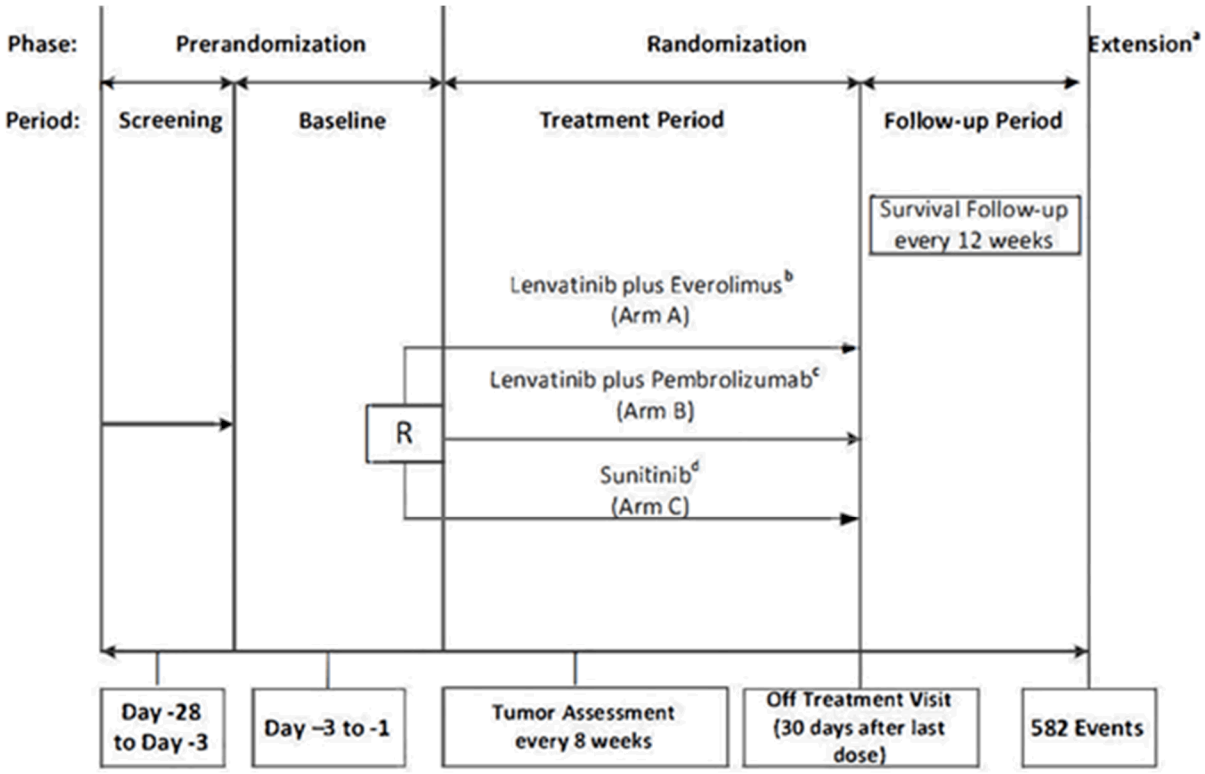

Figure 1 Study Design for Study E7080-G000-307

R = Randomization

a: Extension Phase includes a Treatment and Follow-up Period. All subjects still on treatment at the end of the Randomization Phase will enter the Extension Phase and continue to receive the same study treatment they received in the Randomization Phase.

b: Lenvatinib 18 mg plus everolimus 5 mg given orally once daily.

c: Lenvatinib 20 mg once daily plus pembrolizumab 200 mg intravenously every 3 weeks.

d: Sunitinib 50 mg once daily on a schedule of 4 weeks on treatment followed by 2 weeks off (Schedule 4/2).

The Pre-Randomization Phase will consist of 2 periods: Screening and Baseline. The Pre-Randomization Phase will last no longer than 28 days and will include the Screening Period to establish protocol eligibility and the Baseline Period to confirm eligibility and establish disease characteristics prior to randomization and treatment.

Screening will occur between Day -28 and Day -3. The purpose of the Screening Period is to obtain informed consent and to establish protocol eligibility. The screening assessment can serve as the baseline assessment, if performed within 72 hours before randomization. Informed consent will be obtained after the study has been fully explained to each subject and before the conduct of any screening procedures or assessments.

The purpose of the Baseline Period is to establish disease characteristics prior to treatment and randomization and to confirm protocol eligibility as specified in the inclusion/exclusion criteria. The baseline assessments can be performed within 72 hours prior to randomization on Cycle 1/Day 1. Laboratory tests and a pregnancy test (for female subjects of childbearing potential) may be performed up to 72 hours before randomization. Repeated laboratory evaluation to establish eligibility is not allowed unless discussed and agreed upon with the sponsor. Subjects who complete the Baseline Period and meet the criteria for inclusion/exclusion (Protocol Sections 9.3.1 and 9.3.2) will begin the Randomization/Treatment Phase.

The Randomization Phase will consist of 2 periods: Treatment Period and Follow-up Period. The Randomization Phase will begin at the time of randomization of the first subject and will end on the data cutoff date for the planned final PFS analysis.

The Treatment Period for each subject will begin at the time of randomization and will end with the completion of the Off-Treatment Visit, which will occur within 30 days after the final dose of study treatment.

Subjects will receive study treatment as continuous 21-day cycles. Treatment cycles will be counted continuously regardless of dose interruptions. Subjects will undergo safety and efficacy assessments as defined in the Schedule of Procedures/Assessments. Subjects will continue to receive study treatment until confirmed disease progression by independent review, development of unacceptable toxicity, subject request, withdrawal of consent, completion of 35 treatments (approximately 2 years) with pembrolizumab, or study termination by the sponsor. Subjects in Arm B who discontinue pembrolizumab after 35 treatments may continue treatment with lenvatinib alone unless any other criteria above apply.

Subjects will be permitted to continue study treatment beyond RECIST 1.1-defined progression of the disease as long as the treating investigator considers that there is clinical benefit and, the subject is tolerating study treatment. The assessment of clinical benefit should take into account whether the subject is clinically deteriorating and unlikely to receive further benefit from continued treatment.

All decisions to continue treatment beyond initial progression must be discussed with the Eisai Medical Monitor. Subjects will discontinue study treatment upon evidence of further progression and/or loss of clinical benefit, as judged by the Investigator.

The Follow-up Period will begin the day after the Off-Treatment Visit and will continue as long as the subject is alive, unless the subject withdraws consent or the sponsor terminates the study. If a subject discontinues study treatment and does not consent to continued follow-up, the investigator must not access confidential records that require the subject's consent. However, an investigator may consult public records to establish survival status.

During the follow-up period, subjects will be treated by the investigator according to the prevailing local standard of care. Subjects will be followed every 12 weeks (±1week) for PFS2, survival, and all subsequent anticancer treatments received. This information will be recorded unless this information is not allowed to be provided due to confidentiality. The sponsor may choose to discontinue survival follow-up following completion of the planned final PFS analysis.

All subjects who discontinue study treatment prior to disease progression will continue to undergo tumor assessments every 8 weeks and bone scan every 24 weeks in the Follow-up Period, until disease progression is documented and confirmed by independent review or a new anticancer therapy is initiated, unless the subject withdraws consent or is lost to follow-up.

The end of the Randomization Phase of the study will be the data cutoff date for the primary study analysis, which will occur when approximately 582 PFS events as determined by the IIR among the 3 treatment arms and at least 388 events for Arm A (or Arm B) and Arm C have been observed. All subjects who are still on study treatment at that time will enter the Extension Phase.

The Extension Phase will consist of 2 periods: Treatment Period and Follow-up Period. In the Treatment Period, subjects still on study treatment following the data cutoff date of the planned final PFS analysis (ie, at the end of the Randomization Phase) will continue to receive study treatment in 21-day cycles. Tumor assessments will be performed according to the local standard of care (but not less frequently than every 12 weeks), and scans will no longer be required to be sent to the imaging core lab. Subjects will continue to receive study treatment until disease progression, development of unacceptable toxicity, subject request, withdrawal of consent, loss of clinical benefit, completion of 35 treatments (approximately 2 years) with pembrolizumab, or sponsor termination of the study. Subjects in Arm B who discontinue pembrolizumab after 35 treatments may continue treatment with lenvatinib unless any of the other discontinuation criteria apply. The Off-Treatment Visit will occur within 30 days after the final dose of study treatment.

The Follow-up Period will begin the day after the Off-Treatment Visit and will continue as long as the study subject is alive, unless the subject withdraws consent or the sponsor terminates the study. If a subject discontinues study treatment and does not consent to continued follow-up, the investigator must not access confidential records that require the

subject's consent. However, an investigator may consult public records to establish survival status.

Subjects will be treated by the investigator according to the prevailing local standard of care. Subjects will be followed every 12 weeks (±1week) for PFS2, survival, and all subsequent anticancer treatments received. This information will be recorded unless this information is not allowed to be provided due to confidentiality. The sponsor may decide to terminate survival follow-up after the completion of the primary study analysis.

The definition of the end of the study is the date of the data cutoff for the final overall survival analysis or last subject/last visit/last assessment including discontinuation from study for any reason, whichever occurs later.

4 DETERMINATION OF SAMPLE SIZE

The sample size is estimated based on the primary endpoint PFS. Approximately 1050 subjects will be randomized in a 1:1:1 ratio into one of the three treatment arms: lenvatinib + everolimus, lenvatinib + pembrolizumab, or sunitinib alone. The randomization scheme will be stratified by geographic region (Western Europe and North America versus Other) and MSKCC prognostic groups (favorable, intermediate and poor risk).

The same treatment effect is assumed for the primary comparisons of lenvatinib + everolimus (Arm A) and lenvatinib + pembrolizumab (Arm B) each compared to sunitinib alone (Arm C). Assuming the median PFS in sunitinib alone (Arm C) to be 12.3 months and a hazard ratio of 0.714 for the primary comparisons, this corresponds to 40% improvement (4.9 months) in median PFS from 12.3 months to 17.2 months for Arm A versus Arm C and for Arm B versus Arm C. A yearly loss of PFS event rate of 22% is assumed in the sample size calculation. Since the study is testing more than one comparison for the primary and secondary endpoints, the graphical approach (SAP Section 5.3.3) will be used for testing multiple hypotheses. For the two PFS comparisons (one for each test arm), the sponsor chooses to split the total alpha of 0.0499 (2-sided), as initial allocations, into $\alpha = 0.045$ for the comparison between Arm B and Arm C, and $\alpha = 0.0049$ for the comparison between Arm A and Arm C.

The study is designed to achieve 90% power at $\alpha = 0.045$ to detect a statistically significant difference in PFS in the comparison between Arm B and Arm C. Therefore, a total of 388 PFS events are required between Arms B and C in the final PFS analysis. Since the same number of PFS events are expected to be observed in Arms A and C, a total of 388 PFS events is expected at the final PFS analysis for the comparison between Arms A and C. The power to detect a statistically significant difference in PFS between Arm A and Arm C is approximately 70% at the initial assigned $\alpha = 0.0049$, and will be at least 90% after alpha re-allocation when the hypothesis tests of PFS and OS in the comparison of Arm B and Arm C are statistically significant. In the power calculation for PFS analysis, it is assumed that one interim analysis of PFS is to be performed at the 80% information fraction and a Lan-DeMets spending function with O'Brien-Fleming boundary is used between the interim and final analysis of PFS. The interim analysis of

PFS will be performed when it is approximately 4 months after the last subject is randomized and an approximately 80% information fraction of PFS events (approximately 310 PFS events as determined by the IIR) in Arm B and Arm C. Assuming an average enrolment rate of 31 subjects per month, the interim and final analyses of PFS will occur approximately 38 and 45 months (34 months enrollment period) after first subject is randomized. A total of 582 PFS events are expected in 3 arms by the time of planned final PFS analysis.

For the key secondary endpoint of OS, a total of 304 deaths for each comparison (456 death events among the 3 arms) are expected in the final OS analysis. For OS testing, when the corresponding PFS testing is statistically significant at the initial assigned alpha, the study will provide 80% power to detect a statistically significant difference at an α level of 0.045 for the comparison between Arms B and C, and 50% power at an α level of 0.0049 for the comparison between Arms A and C. By using the graphical approach, the power for the OS comparison between Arms A and C will increase to at least 80% with alpha re-allocation when the OS testing between Arms B and C is significant and both PFS tests are significant. The assumptions that are used for the OS power calculations are: 1) the hazard ratio is 0.70 (median OS is 54.1 months in Arm A or Arm B and 37.9 months in Arm C), 2) interim analyses at approximately 45%, 60%, and 80% information fraction of death events, 3) a Lan-DeMets spending function with Pocock boundary is used, and 4) the yearly rate for loss to follow-up is 3%. With the planned sample size and the assumptions for enrollment, the final analysis of OS is expected to occur approximately 69 months after the first subject is randomly assigned to treatment.

For the key secondary endpoint of ORR, assuming an ORR of 32% in Arm C and 48% in Arm A or Arm B, the study will provide at least 95% power to detect a difference with alpha re-allocation when testing of PFS and OS are positive for each comparison of Arm B vs Arm C and Arm A vs Arm C.

5 STATISTICAL METHODS

All descriptive statistics for continuous variables will be reported using mean, standard deviation (SD), median, Q1, Q3, minimum and maximum, otherwise will be specified. Categorical variables will be summarized as number (percent) of subjects. All summarized variables will also be listed for each subject. Statistical analyses will be performed using SAS software or other validated statistical software as required.

5.1 Study Endpoints

5.1.1 Primary Endpoint

The primary endpoint is PFS assessed by IIR, defined as the time from the date of randomization to the date of the first documentation of disease progression or death (whichever occurs first) using RECIST 1.1 (Eisenhauer et al., 2009). PFS censoring rules will follow the FDA guidance of 2007 (FDA, 2007) and will be detailed in subsequent section.

5.1.2 Secondary Endpoints

- Objective response rate (ORR), defined as the proportion of subjects who have best overall response (BOR) of CR or PR as determined by IIR using RECIST 1.1. ORR will be calculated for confirmed CR/PR, and for confirmed and unconfirmed CR/PR. Confirmed CR/PR will be primary for ORR analysis.
- Overall survival (OS), defined as the time from the date of randomization to the date of death from any cause. Subjects who are lost to follow-up and those who are alive at the date of data cutoff will be censored at the last date the subject was last known alive, or date of data cutoff, whichever occurs first.
- Safety will be assessed summarizing the incidence of treatment-emergent adverse events (TEAEs) and SAEs together with all other safety parameters.
- Proportion of subjects who discontinue treatment due to toxicity is defined as the proportion of subjects who discontinue study treatment due to treatment-emergent adverse events (TEAEs).
- Time to treatment failure is defined as time from the date of randomization to the date that a subject discontinues study treatment due to TEAEs.
- Health-Related Quality of Life (HRQoL), assessed using the Functional Assessment of Cancer Therapy Kidney Syndrome Index-Disease-Related Symptoms (FKSI-DRS), the European Organization for the Research and Treatment of Cancer (EORTC) QLQ-C30 and the European Quality of Life (EuroQOL) EQ-5D-3L instruments. This analysis will be detailed in a separate SAP for HRQoL.
- PFS on next-line therapy (PFS2), defined as the time from randomization to disease progression on next-line treatment or death from any cause (whichever occurs first).
- Progression-free survival (PFS) by investigator assessment is defined as the time from the date of randomization to the date of first documentation of disease

progression based on the investigator assessment per RECIST v.1.1 or death (whichever occurs first).

- Pembrolizumab PK comparison to historical data. This analysis will be detailed in a separate SAP for PK/PD.
- Model-predicted clearance and AUC for lenvatinib in Arms A and B. This analysis will be detailed in a separate SAP for PK/PD.
- Model-predicted clearance and AUC for everolimus in Arm A. This analysis will be detailed in a separate SAP for PK/PD.

5.1.3 Exploratory Endpoints

Exploratory efficacy endpoints will be:

- Objective response rate (ORR), defined as the proportion of subjects who have best overall response (BOR) of CR or PR as determined by investigator assessment using RECIST 1.1.
- Duration of response (DOR) is defined as the time from the date a response of CR or PR by IIR and investigator assessment was first documented until the date of the first documentation of disease progression or date of death from any case.
- Disease control rate is the proportion of subjects who have best overall response of CR or PR or SD by IIR and investigator assessment. Stable disease must be achieved at ≥ 7 weeks after randomization to be considered best overall response.
- Clinical benefit rate is the proportion of subjects who have best overall response of CR or PR or durable SD (duration of SD ≥ 23 weeks after randomization) by IIR and investigator assessment.
- Blood and tumor biomarkers will be assessed for identifying potential correlation with clinical outcomes-related endpoints. This analysis will be detailed in a separate SAP for biomarkers.

For exploratory efficacy endpoints that involve with CR/PR, the endpoint will be calculated once for confirmed CR/PR, and once for confirmed and unconfirmed CR/PR.

5.2 Study Subjects

5.2.1 Definitions of Analysis Sets

The following analysis sets will be defined:

- The Full Analysis Set (Intent-to-Treat Analysis [ITT] Population) is the group of all randomized subjects regardless of the treatment actually received. This is the primary analysis population used for all efficacy analyses which will be based on the intent-to-treat principle.
- The Per Protocol Analysis Set is the subjects who received at least one dose of

any study drug, had no major protocol deviations and had both baseline and at least 1 postbaseline tumor assessments. Subjects who died prior to the first postbaseline tumor assessment will also be included. The per protocol analysis set will be the secondary analysis set for efficacy endpoints.

- The Safety Analysis Set is the group of subjects who received at least one dose of any study drug. This is the analysis population for all safety analyses, which will be based on as-treated principle.
- Population Pharmacokinetic (PK) Analysis Set: All subjects who have received at least one dose of study treatment with documented dosing history in the lenvatinib plus everolimus arm (Arm A) or the lenvatinib plus pembrolizumab arm (Arm B), and have measurable plasma levels of lenvatinib or whole blood levels of everolimus.

Table 1 Analysis Sets in Data Analysis

Data Category	Full	Per Protocol	Safety
Protocol Deviations	•		
Disposition	•		
Demography & Baseline Characteristics	•		
Disease History	•		
Prior & Concomitant Medications	•		
Progression-Free Survival	•	•	
Tumor Response	•	•	
Overall Survival	•	•	
Drug Exposure			•
Adverse Events			•
Deaths			•
Laboratory Tests			•
Vital Signs			•
ECGs			•
Echocardiography			•

- Pembrolizumab PK Analysis Set: All subjects who have received at least one dose of study treatment with documented dosing history in the lenvatinib plus pembrolizumab arm (Arm B) and have measurable serum concentrations of pembrolizumab.

- The Pharmacodynamic Analysis Set is the group of subjects who received at least one dose of study drug and had sufficient pharmacodynamic data to derive at least one pharmacodynamic parameter and with documented dosing history.
- The HRQoL Analysis Set will consist of all subjects who have any HRQoL data.

Table 1 shows each analysis set to be primarily used in each data category but not limited to the specifications in this table.

5.2.2 Subject Disposition

Enrolled subjects include all subjects who signed informed consent forms. The number of subjects screened and the number (percent) of subjects who failed screening and the reasons for screen failure will be summarized, based on data reported on the Screening Disposition Case Report Form (CRF).

The number of subjects randomized, treated, prematurely discontinued from study drug, and those with major protocol deviations will be counted by site and by treatment arms. The number of subjects who discontinued one medication but remained on other medication in Arm A and Arm B will also be summarized. The primary reason for study drug discontinuation will be summarized by treatment arms and overall according to the categories in the corresponding CRFs.

The end of study status: alive (still on treatment or on follow up), death, withdrew consent, lost to follow-up or study terminated by sponsor at the data cutoff date will also be summarized by treatment arms and overall.

5.2.3 Protocol Deviations

Major protocol deviation criteria will be established and subjects with major protocol deviations will be identified and documented before the database lock for the primary analysis. Major protocol deviations will be appropriately summarized by center and grouped into different categories. All major protocol deviations identified according to study entry criteria and during treatment will be listed, broken down by center.

5.2.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics for the Full Analysis Set will be summarized for each treatment arm and overall using descriptive statistics for the following variables. They include but are not limited to age, sex, race, race group, ethnicity, weight, height, body mass index (BMI), geographic region, Karnofsky performance status (KPS), KPS score group, MSKCC prognostic group, International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) prognostic group and PD-L1 status. Disease history and characteristics at study entry, previous anticancer medications and previous radiotherapies will be summarized. Palliative radiotherapy on treatment for RCC will be listed separately.

Subject listings of medical history and non-pharmacological procedures on treatment will be provided.

5.2.5 Prior and Concomitant Therapy

All investigator terms for medications recorded in the CRF will be coded to an 11-digit code using the World Health Organization Drug Dictionary (WHO DD) version of March 2018 or later. The number (percentage) of subjects who took prior and concomitant antihypertensive, antidiarrheal medications and corticosteroids for systemic use will be summarized on the Full Analysis Set by treatment arm, Anatomical Therapeutic Chemical (ATC) class 1, pharmacological class 3, and WHO DD preferred term. Other prior and concomitant medications, excluding antihypertensive, antidiarrheal medications and corticosteroids for systemic use will be summarized similarly. If a subject has taken prior and concomitant medications more than once within the same preferred term, the subject will be counted only once for the respective preferred term. Prior medications are defined as medications that started prior to the first dose of study drug. Concomitant medications are defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 30 days after the subject's last dose. A medication that cannot be determined as prior/concomitant/post treatment due to missing/incomplete dates will be regarded as a concomitant medication. All medications will be presented in subject data listings.

5.3 Data Analysis General Considerations

Tumor response will be assessed by IIR and the investigator based on RECIST 1.1 for all three arms. Analyses will be performed separately for assessments by IIR and by the investigator, but the primary endpoint is progression-free survival as assessed by IIR.

All stratified efficacy analyses will be based on strata recorded in the IxRS system at randomization as primary. The concordance and discordance of strata between IxRS and the clinical database will be summarized in frequency and percentage by treatment arms and listed by treatment arm and subject.

5.3.1 Pooling of Centers

This study is a multicenter, international study. Subjects from all centers for each region will be pooled and region will be used as one of the two stratification factors for most of efficacy analyses.

5.3.2 Adjustments for Covariates

Cox proportional hazards models and log-rank tests for PFS and OS, and the Cochran-Mantel-Haenszel (CMH) test of ORR will be stratified by region (West Europe and North America, Rest of World), and MSKCC Risk Group (favorable, intermediate, and poor risk). Analyses of the primary endpoint PFS will also include other covariates of interest by means of subgroups.

5.3.3 Multiple Comparisons/Multiplicity

To adjust for multiplicity and provide strong control of the overall family-wise error rate (FWER), the graphical approach of Maurer and Bretz (Maurer et al., 2013) will be used in the primary endpoint of PFS and the key secondary efficacy endpoints (OS and ORR). No multiplicity adjustment will be made for other secondary endpoints analyses and exploratory endpoints analyses. Subgroup analyses in Section 5.3.4 will be considered as exploratory analyses and no multiplicity adjustment will be made.

An α of 0.0001 will be subtracted from the total α of 0.05 to account for the interim analysis of ORR from Arm B. Figure 2 shows the initial α-allocation (the remaining α = 0.0499) for each hypothesis and the graphical approach for multiple analyses of PFS, OS, and ORR.

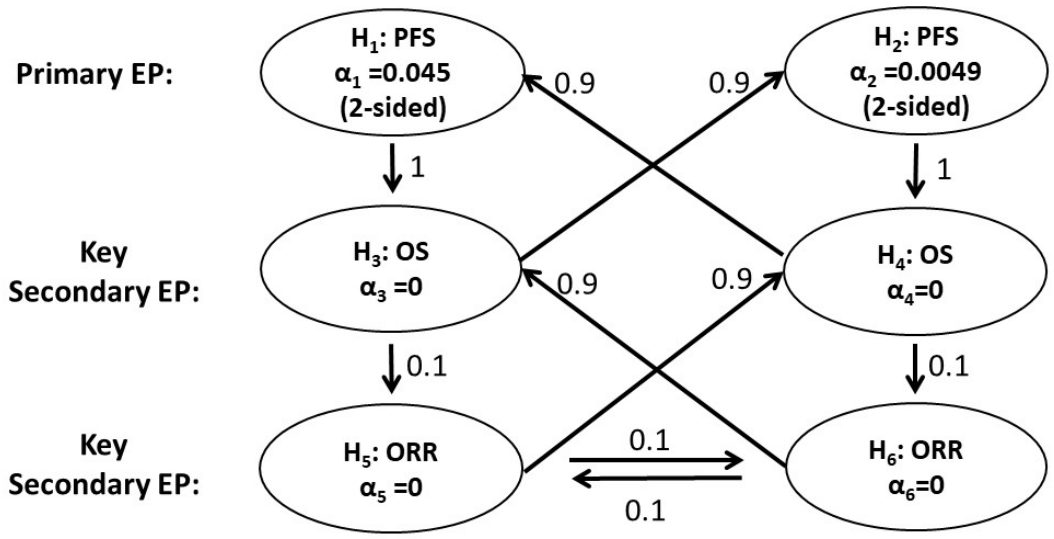

Figure 2 Graphical Approach to Control Familywise Error Rate for Testing Primary and Key Secondary Endpoints

EP = endpoint; ORR = objective response rate; OS = overall survival; PFS = progression-free survival.
Hypothesis (H_1): The PFS of lenvatinib + pembrolizumab arm is superior to that of sunitinib arm.
Hypothesis (H_2): The PFS of lenvatinib + everolimus arm is superior to that of sunitinib arm.
Hypothesis (H_3): The OS of lenvatinib + pembrolizumab arm is superior to that of sunitinib arm.
Hypothesis (H_4): The OS of lenvatinib + everolimus is superior to that of sunitinib arm.
Hypothesis (H_5): The ORR of lenvatinib + pembrolizumab arm is superior to that of sunitinib arm.
Hypothesis (H_6): The ORR of lenvatinib + everolimus is superior to that of sunitinib arm.

Table 2 Efficacy Boundaries and Properties for PFS H_1 and PFS H_2 (LDOF spending function) Based on Initial Assigned Alpha

Analysis (2 arms)	Value	H_1 (α=0.045)	H_2 (α=0.0049)[d]
IA: 80%[a] N: 700 Events: 310 Month: 38	*P* (2-sided)[b]	0.0216	0.0014
	HR at boundary[c]	0.7705	0.6964
	Power	75%	42%
Final: 100% N: 700 Events: 388 Month: 45	*P* (2-sided)[b]	0.0386	0.0046
	HR at boundary[c]	0.8105	0.7491
	Power	90%	69%

HR = hazard ratio, IA = interim analysis, N = number of subjects.
a: Information fraction, percentage of expected number of events at final analysis.
b: P-value (2-sided) is the nominal α for testing.
c: HR at boundary is the approximate HR required to reach an efficacy boundary.
d: The power of H_2 test will be at least 90% if H_1 and H_3 testing are significant.

Table 3 Efficacy Boundaries and Properties for OS H_3 and OS H_4 (LD-Pocock Spending function) when PFS Tests Are Significant

Analysis (2 arms)	Value	H_3 (α=0.045)	H_4 (α=0.0049)[d]
IA: 45%[a] N: 700 Events: 137 Month: 38	*P* (2-sided)[b]	0.0258	0.0028
	HR at boundary[c]	0.683	0.600
	Power	44%	20%
IA: 60%[a] N: 700 Events: 182 Month: 45	*P* (2-sided)[b]	0.0152	0.0014
	HR at boundary[c]	0.698	0.622
	Power	55%	27%
IA: 80%[a] N: 700 Events: 243 Month: 57	*P* (2-sided)[b]	0.0158	0.0014
	HR at boundary[c]	0.734	0.663
	Power	69%	41%
Final: N: 700 Events: 304 Month: 69	*P* (2-sided)[b]	0.0158	0.0014
	HR at boundary[c]	0.758	0.692
	Power	80%	51%

HR = hazard ratio, IA = interim analysis, N = number of subjects.
a: Information fraction, percentage of expected number of events at final analysis.
b: P-value (2-sided) is the nominal α for testing.
c: HR at boundary is the approximate HR required to reach an efficacy boundary.
d: The power of H_4 test will be at least 80% if H_1, H_3 and H_2 are significant.

The nominal α level for each PFS comparison at the interim and final analyses will be determined by a Lan-DeMets spending function with an O'Brien-Fleming boundary as illustrated in Table 2. The nominal α level for each OS comparison at the interim and final analyses will be determined by a Lan-DeMets spending function with Pocock boundary as illustrated in Table 3. The actual boundaries will be calculated using the observed number of events at the interim and final analyses and α passed from previous tests.

5.3.4 Examination of Subgroups

For efficacy endpoints, the hazard ratio and two-sided 95% confidence interval (CI) for comparing PFS as assessed by the IIR and investigator assessment of either Arm A versus Arm C or Arm B versus Arm C will be presented in forest plots for the subgroups. Median PFS and 95% CIs will be presented for all subgroups. Similar summary and plots will be provided for OS. In addition, the odds ratio and two-sided 95% CIs for comparing ORR as assessed by IIR will be summarized and presented in forest plots. If sample size of subgroup/strata is less than 18 (5% of sample size in the treatment group), the subgroup/strata may be considered to collapse to the closest subgroup/strata as appropriate.

- Age group (<65 years, ≥65 years)
- Sex (male, female)
- Race (White, Asian, all others)
- Geographic region per IxRS (West Europe and North America, Rest of World)
- MSKCC risk group per IxRS (favorable , intermediate, poor risk)
- IMDC risk group (favorable , intermediate, poor risk)
- Number of metastatic sites per IIR (0, 1, 2, ≥3)
- KPS score group (100-90, 80-70)
- Baseline bone metastasis (yes, no)
- Baseline liver metastasis (yes, no)
- Baseline lung metastasis (yes, no)
- PD-L1 status (CPS≥1, CPS<1, not available)
- Prior nephrectomy (yes, no)
- Histologic clear component featuring sarcomatoid (yes, no)

Additional subgroup analyses may also be conducted, if deemed appropriate.

5.3.5 Handling of Missing Data, Dropouts, and Outliers

Adverse Events with incomplete start dates will be considered treatment emergent if:

a) Day and month are missing and the year is equal to or after the year of the first dose date;
b) Day is missing, and the year is after the year of the first dose;
c) Day is missing and the year is equal to the year of the first dose date and the month is equal to or after the month of the first dose date;
d) Year is missing; or
e) Complete date is missing.

Medications will be considered concomitant if:

a) Day and month are missing and the year is equal to or after the year of the first dose date;
b) Day is missing, and the year is after the year of the first dose;
c) Day is missing and the year is equal to the year of the first dose date and the month is equal to or after the month of the first dose date; or
d) Year is missing; or
e) Complete date is missing.

For incomplete dates involving efficacy and other safety data, a conservative imputation will be used for calculation if needed. More details of imputation rules will be specified in study analysis dataset specification.

5.3.6 Other Considerations

In case subjects who took combination therapy in Arm A and Arm B discontinue one treatment earlier than the other, end of treatment date is defined as the latest date of the last dose dates of two study drugs. All efficacy and safety endpoints will be summarized up to the end of treatment date plus applicable window for each endpoint.

5.4 Efficacy Analyses

5.4.1 Primary Efficacy Analyses

The primary endpoint PFS as assessed by IIR using RECIST 1.1 will be compared between lenvatinib + everolimus (Arm A) and sunitinib (Arm C), as well as between lenvatinib + pembrolizumab (Arm B) and sunitinib (Arm C). An interim and a final analysis of PFS are planned to be performed (SAP Section 6). Lan-DeMets spending function with O'Brien-Fleming boundary will be used to control alpha levels between the interim and final analysis of PFS (SAP Section 5.3.3). The interim analysis of PFS will be performed when it is approximately 4 months after the last subject is randomized and an approximately 80% information fraction of PFS events (as determined by the IIR) in Arm B and Arm C. The final analysis of PFS will be performed when approximately 388 PFS events, as determined by the IIR, are observed between each comparison. A graphical approach will be used to control the FWER at the two-sided 0.0499 for multiple comparisons, including both PFS, OS and ORR comparisons of Arm B vs Arm C and Arm A vs Arm C. For each comparison, a statistical significance can be claimed based on either interim or final analysis of PFS at specified alpha levels.

The PFS curve in each treatment group will be estimated using Kaplan-Meier (K-M) method and the difference in PFS for the 2 primary comparisons will each be tested by stratified log-rank test. The tests will be stratified by geographic region and MSKCC prognostic groups. The hazard ratio (lenvatinib + everolimus relative to sunitinib and lenvatinib + pembrolizumab relative to sunitinib) and the corresponding 95% CIs will be estimated using the Cox regression model with Efron's method for ties, stratified by the factors used for stratified randomization. The median PFS, and the PFS rates at 6, 12, 18 months, and so on (depending on data adequacy), will be calculated using the K-M product-limit estimates for each treatment arm, and presented with two-sided 95% CIs.

The K-M curve of PFS will be plotted over time.

All primary efficacy analyses will be performed based on Full Analysis Set as the primary analysis set and efficacy analyses performed on the Per Protocol Analysis Set will be considered as the secondary analysis set.

The PFS censoring rules in this SAP (Table 4) and definition of progression date follow the Food and Drug Administration (FDA) "Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (2007)" and European Medicines Agency (EMA) "Guideline on the Evaluation of Anticancer Medicinal Products in Man (2013)".

Progression date is assigned to the earliest date when any RECIST 1.1-defined disease progression is observed without missing more than one adequate radiologic assessment.

Table 4 Censoring Rules for Derivation of Progression-Free Survival

No.	Situation	Date of Progression or Censoring	Outcome
1	No baseline or postbaseline tumor assessments	Date of randomization	Censored
2	Progression documented between scheduled visits	Date of first radiologic PD assessment	Progressed
3	No progression at the time of data cutoff	Date of last adequate radiologic assessment prior to or on date of data cutoff	Censored
4	New anticancer treatment started	Date of last adequate radiologic assessment prior to or on date of new anticancer treatment	Censored
5	Death before first PD assessment	Date of death	Progressed
6	Death between adequate assessment*	Date of death	Progressed
7	Death or progression after more than one missed visit/tumor assessment**	Date of last adequate radiologic assessment before missed tumor assessments	Censored

CR = complete response, PD = progressive disease, PR = partial response, SD = stable disease.

* Adequate tumor assessment is radiologic assessment of CR, PR, SD, non-CR/non-PD or PD as determined by investigators at regular interval as defined in the protocol. Any tumor assessments after new anticancer treatment starts will be removed in the definition of PFS.

** More than one missed visit/adequate tumor assessment is defined as having the duration between the last adequate tumor assessment and PD or death being longer than 16 weeks + 10 days (tumor assessment window) - 1 day, which is 121 days for subjects on the every 8 week tumor assessment schedule in this study.

The priority of the censoring rules is as follows:

1. If the subject had PD or death, the following sequence will be applied:

- If a subject did not have a baseline tumor assessment (No. 1), the subject will be censored on the date of randomization. However, if the subject died within 121 days after randomization and did not receive a new anticancer treatment, it will be counted as PFS event at the date of death. If a subject had new anticancer treatment before PD or death (No. 4), the subject will be censored on the date of the last adequate tumor assessment prior to or on the date of new anticancer treatment.
- If a subject missed two or more tumor assessments before PD or death (No. 7), the subject will be censored on the date of the last adequate tumor assessment before PD or death. Note that if a subject is censored by both this criterion and the anticancer treatment criterion, the earliest censoring date will be used.
- Otherwise, if a subject had an event (No. 2, No. 5, or No. 6), the earliest event date will be used.

2. If a subject did not have PD or death, the censoring date will be the earliest censoring date if the subject met multiple censoring criteria (No. 1, No. 3, No. 4, No. 7).

Sensitivity analyses will be performed using unstratified log rank tests for comparisons of PFS of lenvatinib + everolimus (Arm A) versus sunitinib alone (Arm C) and lenvatinib + pembrolizumab (Arm B) versus sunitinib alone (Arm C), as well as unstratified Cox proportional hazards model with Efron's method for ties, including treatment arms as a single covariate for the estimation of the hazard ratio. In addition, the following sensitivity analyses will also be performed:

1. Using the actual reported date of progression by IIR or death to define PFS regardless of missing assessments, or use of new anti-cancer therapy (Per EMA guidance);
2. Using the radiologic assessment data as assessed by Investigator and death to define PFS;
3. Using the different derivation rule for the situation with more than one missed visit/tumor assessment: death or progression immediately after more than one missed visit/tumor assessment (i.e. if a subject missed two or more tumor assessments right before PD or death, the subject will be censored on the date of the last adequate tumor assessment before PD or death. Note that if a subject is censored by both this criterion and the anticancer treatment criterion, the earliest censoring date will be used).

5.4.2 Secondary Efficacy Analyses

All analyses of secondary efficacy endpoints will be performed based on Full Analysis Set as the primary analysis set and Per Protocol analysis set as the secondary analysis set.

1. Overall Survival (OS) will be compared between lenvatinib + everolimus (Arm A) and sunitinib alone (Arm C), and between lenvatinib + pembrolizumab (Arm B) and sunitinib alone (Arm C), using the stratified log-rank test with geographic region and MSKCC prognostic groups as strata. The hazard ratio and its 95% CI comparing lenvatinib + everolimus (Arm A) versus sunitinib alone (Arm C) and lenvatinib + pembrolizumab (Arm B) versus sunitinib alone (Arm C) will be estimated by a stratified Cox proportional hazards model with Efron's method for ties, stratified by geographic region and MSKCC prognostic groups. Median OS

and the OS rate at 12, 18, 24 months, and so on (depending on data adequacy) with 2-sided 95% CIs will be calculated using K-M product-limit estimates for each treatment arm and K-M estimates of OS will be plotted over time. Three interim and a final analyses of OS are planned to be performed (SAP Section 6). Lan-DeMets spending function with Pocock boundary will be used to control alpha levels among the interim and final analysis of OS (SAP Section 5.3.3). The first two OS interim analyses will be performed at the time of PFS interim and final analysis, corresponding to approximately 45% and 60% of information fractions of OS events. The third OS interim analysis will be performed at approximately 80% information fraction of OS events. The final analysis of OS will be performed when 304 OS events are observed for each comparison.

All events of death will be included, regardless of whether the event occurred while the subject was still taking study drug, or after the subject discontinued study drug. Subjects who were alive at the data cutoff date will be censored at the data cutoff date. Subjects who were lost to follow-up or withdrew consent will be censored at the date the subject was last known to be alive.

2. Objective Response Rate (ORR), estimated by treatment arms based on the tumor response evaluation by IIR according to RECIST 1.1, will be calculated with exact 95% confidence intervals using the method of Clopper and Pearson. The difference between treatment arms comparing lenvatinib + everolimus (Arm A) versus sunitinib alone (Arm C) and lenvatinib + pembrolizumab (Arm B) versus sunitinib alone (Arm C) will be tested using the Cochran-Mantel-Haenszel (CMH) test, stratified by geographic region and MSKCC prognostic groups. The 2-sided 95% CIs for the odds ratio and the difference in ORR will be calculated. The p-value for hypothesis testing of ORR will be based on the ORR data at the time of the PFS interim analysis. The ORR data available at the subsequent analysis time points will be provided for supportive purposes.

For the adjustment of multiplicity of primary endpoint PFS, key secondary endpoints ORR and OS, please refer to SAP Section 5.3.3. There is no multiplicity adjustment to be made for other secondary and exploratory efficacy endpoints. For statistical tests to be performed on other secondary and exploratory efficacy endpoints, nominal p-values will be generated.

PFS2 will be evaluated using K-M estimates and difference in PFS2 will be tested by stratified log-rank test, stratified by geographic region and MSKCC prognostic groups. The hazard ratio (lenvatinib + everolimus relative to sunitinib and lenvatinib + pembrolizumab relative to sunitinib) and the corresponding 95% confidence intervals (CIs) will be estimated using the Cox regression model with Efron's method for ties, stratified by the same stratification factors. Patients alive and for whom a disease progression on second line therapy has not been observed will be censored at the last time known to be alive. Subjects who did not have second line therapy and no death documented before data cutoff date will be censored at the data cutoff date. Time to next-line therapy will be summarized only for subjects who experience the next-line therapy. It is defined as the time between

randomization and the start date of next-line therapy. The median and quantile of time to initiation of next-line therapy will be summarized.

5.4.3 Other Efficacy Analyses

All statistical tests specified in this section will be performed for exploratory purpose and no multiplicity adjustment will be made.

5.4.3.1 Other Analyses of Tumor Response Using RECIST 1.1

The best overall response categories (CR, PR, SD [including non-CR/non-PD], durable SD (duration of SD ≥ 23 weeks), PD, NE, Unknown) will be derived based upon time point tumor responses during the study as assessed by IIR as well as the Investigator. BOR of SD must occur at least 49 days after randomization. If a subject has a BOR of non-CR/non- PD, the subject's BOR will be grouped with the SD category.

The counts/percentages will be calculated for each BOR category by treatment group. Concordance and discordance on BOR for each subject between assessment by IIR and by investigator review will be analyzed. Also, Cohen's kappa statistic will assess the overall concordance for each treatment group (Fleiss et al., 2003).

Median duration of response among responders for each treatment arm will be presented along with its corresponding 2-sided 95% CIs. Censoring rules for DOR are the same as those of PFS (Table 4).

Disease Control Rate (DCR) and Clinical Benefit Rate (CBR) will be calculated with exact 95% confidence intervals using the method of Clopper and Pearson. The differences and odds ratios of the above rates between treatment arms and corresponding two-sided 95% CIs will be calculated respectively. These analyses will be performed based on both IIR and investigator assessments. The null hypothesis of no difference in DCR and CBR comparing lenvatinib + everolimus (Arm A) versus sunitinib alone (Arm C) and lenvatinib + pembrolizumab (Arm B) versus sunitinib alone (Arm C) will be tested using the CMH test stratified by geographic region and MSKCC prognostic groups.

The difference in PFS between lenvatinib in combination with everolimus (Arm A) versus lenvatinib in combination with pembrolizumab (Arm B) will be tested by stratified log-rank test, stratified by geographic region and MSKCC prognostic groups. The hazard ratio and the corresponding 95% CI will be estimated using the Cox regression model with Efron's method for ties, stratified by the same stratification factors. The difference in OS between lenvatinib in combination with everolimus (Arm A) versus lenvatinib in combination with pembrolizumab (Arm B) will be analyzed similarly.

Exploratory analyses may also be conducted using IMDC risk category and geographic region as the stratification factors for the endpoints PFS and OS.

An analysis of overall concordance on PFS for each treatment group between assessment by IIR and by investigator review will be performed.

In addition, exploratory analyses may be conducted to include geographic region, MSKCC prognostic groups and other subgroups as covariates, respectively, for the endpoints PFS, OS, and PFS2.

5.5 Pharmacokinetic, Pharmacodynamic, Pharmacogenomic, and Other Biomarker Analyses

PK sampling was to be performed for all subjects of serum lenvatinib (Arms A and B), and whole blood everolimus (Arm A). Details of the analysis methods for population PK/PD modeling will not be described in this SAP but will be described in a separate analysis plan.

5.5.1 Pharmacokinetic Analyses

Scatter plots of dose normalized plasma concentration and associated tabulations using n, mean, SD, percent coefficient of variation (% CV), geometric mean, median, minimum, and maximum will be generated by actual sample collection time and for each PK sample collection visit (Cycle 1 Day 1, Cycle 1 Day 15 and Cycle 2 Day 1). All PK data, which includes the complete bioanalytical results and time and date of blood draws and study drug administration recorded on the CRF, will be presented in subject listings. Plasma concentrations of lenvatinib and whole blood concentrations of everolimus versus time data will be analyzed using a population PK approach to estimate population PK parameters for each respective drug. For lenvatinib, data from this study will be pooled with historical data from other Phase 1 and 2 studies. For everolimus, data from this study will be pooled with data from Study E7080-G000-205. While pembrolizumab concentrations will be compared graphically to historical data. Analysis of serum antidrug antibodies (ADA) and neutralizing antibodies (NAb) to pembrolizumab will be described in a stand-alone report.

5.5.2 Pharmacodynamic, Pharmacogenomic, and Other Biomarker Analyses

Soluble, tissue, genetic and/or imaging biomarkers (baseline and/or posttreatment) may be summarized using descriptive statistics. The correlation between biomarkers and clinical outcomes-related endpoints for safety and/or efficacy (including best overall response, PFS and OS) will be explored as appropriate. Details will be included in a separate analysis plan.

5.5.3 Pharmacokinetic-Pharmacodynamic Analyses

The effect of lenvatinib in combination with everolimus or pembrolizumab on soluble, tissue, genetic and/or imaging biomarkers may be summarized using descriptive statistics in the PK/pharmacodynamic analysis set. PK/pharmacodynamic relationships (i.e., exposure-efficacy, and exposure-safety and exposure-biomarker relationships) may be explored for effects of study treatment. Efficacy endpoints will include the primary endpoint of PFS and other efficacy-related metrics including but not limited to ORR (based on RECIST 1.1) and OS. Safety endpoints will include most frequent AEs of special interest and dose reductions. Exploratory/graphical analyses will be conducted for PK/pharmacodynamic evaluations, and may be followed by model-based analyses. A detailed analysis plan will be provided separately later.

5.6 Safety Analyses

All safety analyses will be performed on the Safety Analysis Set. Safety data, presented by treatment groups will be summarized on an "as treated" basis using descriptive statistics (eg, n, mean, standard deviation, median, Q1, Q3, minimum, and maximum for continuous variables; n [%] for categorical variables). Safety variables include treatment-emergent adverse events (TEAEs), clinical laboratory parameters, vital signs, 12-lead ECG and echocardiogram results including left ventricular ejection fraction (LVEF). Study Day 1 for all safety analyses is defined as the earliest date of the first dose of study drugs.

5.6.1 Extent of Exposure

5.6.1.1 Study Treatment Extent of Exposure

Duration of treatment for study medications individually and/or in combination (overall) will be summarized with descriptive statistics. The dosing end date will be imputed to the analysis cutoff date if the subject is still on treatment at the time of the data cutoff, and the dose will be imputed with the last dose recorded in the database for that subject.

The duration of individual study drug in months will be (Date of the last dose – Date of the first dose + 1)/30.4375, including drug interruption days. For overall treatment duration, it is defined as the duration between the earliest first dose start date of study drugs and the latest last dose end date of study drugs.

5.6.1.2 Study Drug Administration

Percentage of planned dose received, total dose, and dose intensity will be summarized with descriptive statistics. In addition, number of administrations will be presented for pembrolizumab.

The total dose per subject (mg) will be calculated as the sum of all doses per subject. The dose intensity (mg/day) for lenvatinib, everolimus and sunitinib will be calculated as: Total dose (mg) / Duration of individual drug treatment (days).

5.6.1.3 Study Drug Dose Reductions and Interruptions

Reasons for dose reduction and interruption, time to the first dose reduction and time to the first dose interruption for lenvatinib, everolimus and sunitinib will be summarized. Dose reduction refers to the situation that a dose level was reduced from the previous dose level without going back.

In the combination of lenvatinib plus everolimus treatment arm, the starting dose of lenvatinib is 18 mg/day. Lenvatinib dose reductions occur in succession based on the previous dose level (14, 10, and 8 mg/day). The starting dose of everolimus is 5 mg every day. Everolimus dose may reduce to 5 mg every other day until everolimus discontinue. In the combination of lenvatinib plus pembrolizumab treatment arm, the starting dose of lenvatinib is 20 mg/day. Lenvatinib dose reductions occur in succession based on the previous dose level (14, 10, and 8 mg/day and so forth). The starting dose of pembrolizumab is 200 mg per administration. Pembrolizumab must be withheld/interrupted or discontinued for drug related toxicities and severe or life-threatening AEs. In sunitinib treatment arm, the starting dose of sunitinib is

50 mg/day on a Schedule 4/2 for subjects enrolled in Arm C. Sunitinib dose reductions occur in succession based on the previous dose level (37.5, 25 mg/day schedule 4/2).

Dose reduction could be applied to lenvatinib (either 20mg/day or 18mg/day reduce to 14, 10, and 8 mg/day), everolimus (5 mg every day reduce to 5 mg every other day), and sunitinib (50 mg/day reduce to 37.5, 25 mg/day of schedule 4/2). Dose interruption could apply to all study drugs. Definition of dose interruption only includes the scenario that the before and after dose 0 (interruption period), the dose levels or dosing frequency are the same. For example: 20 mg lenvatinib followed by 0 mg and followed by 20 mg lenvatinib; 5 mg everolimus every day followed by 0 mg followed by 5 mg everolimus every day. If dose level reduces from previous dose level after dose interruption period (dose 0), it should count into dose reduction not dose interruption. For example: (1) 20 mg/day lenvatinib followed by 0 mg followed by 14 mg/day lenvatinib; (2) 5 mg everolimus every day followed by 0 mg followed by 5 mg everolimus every other day; the period with 0 mg should not count into dose interruption instead should count as dose reduction. If after dose 0 mg, the subject discontinued from treatment permanently, it should count as treatment discontinuation instead of dose interruption.

Number of subjects with dose reductions and dose interruptions of lenvatinib, everolimus, sunitinib, and pembrolizumab (dose reduction not applied to pembrolizumab) will be summarized by frequency counts and percentages according to study medication data.

Time to first dose reduction is defined as the date of the first dose to the date of first dose reduction, or the date of first dose interruption if subjects had dose interruption followed by dose reduction. There will be two ways to estimate the time to dose reduction. (1) Only include the subjects who had dose reduction events. Descriptive statistics will be generated to describe the time to first dose reduction for these patients. (2) Include all subjects. For subjects who did not have dose reduction, it will be censored at the date of last dose. For this summary, Kaplan-Meier method will be used to estimate the median time to first dose reduction for all the subjects by the treatment arms.

5.6.2 Adverse Events

The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to the MedDRA (Version 21.1 or higher) lower level term (LLT) closest to the verbatim term. The linked MedDRA preferred term (PT) and primary system organ class (SOC) will also be captured in the database. The Common Terminology Criteria for Adverse Events (CTCAE) version 4.03 will be used to assess the severity of AEs.

A treatment-emergent adverse event (TEAE) is defined as an AE that emerged during treatment (up to 30 days after the subject's last dose of study drug), having been absent at pretreatment (Baseline) or

- Reemerged during treatment or up to 30 days following last dose of study drug, having been present at pretreatment (Baseline) but stopped before treatment, or
- Worsened in severity during treatment or up to 30 days following last dose of study drug relative to the pretreatment state, when the AE was continuous.

The serious AE that emerged up to 120 days after the subject's last dose of study drug is also counted as TEAE.

Only those AEs that are treatment emergent will be included in summary tables. All AEs, treatment emergent or otherwise, will be presented in subject data listings.

An overview table, including the incidence of and the number of subjects with TEAEs, TEAEs with grade 3 or above, serious adverse events (SAEs), deaths, and TEAEs that led to treatment discontinuation (discontinuation of both agents in arm A or B), dose reduction, or dose interruption will be provided by treatment arms.

The TEAEs will be summarized by treatment groups. The incidence of TEAEs will be reported as the number (percentage) of subjects with TEAEs by SOC and PT. A subject will be counted only once within a SOC and PT, even if the subject experienced more than 1 TEAE within a specific SOC and PT. The number (percentage) of subjects with TEAEs will also be summarized by maximum severity.

The number (percentage) of subjects with TEAEs will also be summarized by relationship to study drug. Summaries of TEAE incidence rate adjusted by treatment duration will be presented.

In summary, the following TEAE tables will be provided:

- Overview of TEAE and treatment-related TEAE
- TEAE by SOC and PT (Any Grade and Grade ≥ 3, maximum grade 3, 4 and 5)
- TEAE by decreasing frequency of PT
- TEAE with each PT frequency ≥ 5% by PT
- TEAE by SOC, PT and Worst Grade of Toxicity
- Treatment-related TEAE by SOC and PT (Any Grade and Grade ≥ 3)
- Exposure-adjusted TEAEs

The following subject AE listings (treatment emergent or otherwise) will be provided:

- All AEs
- All AEs with CTCAE Grade 3 or above

The proportion of subjects who discontinue treatment due to toxicity will be summarized by frequency counts and percentages. Median, upper and lower quartiles of time to treatment failure (discontinuation of both agents in arm A or B) due to toxicity will be summarized for subjects who discontinue study treatment due to TEAEs.

5.6.2.1 Deaths, Serious and Other Significant Adverse Events

The number (percentage) of subjects with TEAEs leading to death will be summarized by MedDRA SOC and PT for each treatment group. A subject data listing of all AEs leading to death will be provided.

The number (percentage) of subjects with treatment-emergent serious adverse events (SAEs) was summarized by MedDRA SOC and PT for each treatment group. A subject data listing of all SAEs was provided.

The number (percentage) of subjects with TEAEs leading to discontinuation from study drug will be summarized by MedDRA SOC and PT for each treatment group. A subject data listing of all AEs leading to discontinuation from study drug will be provided. The number (percentage) of subjects with TEAE leading to study drug dose reduction or interruption will also be summarized by MedDRA SOC and PT for each treatment group. A subject data listing of all AEs leading to study drug dose reduction or interruption will be provided.

In summary, the following TEAE tables will be provided:

- TEAE leading to death by SOC and PT
- SAEs by SOC and PT
- Treatment-related SAEs by SOC and PT
- SAEs by decreasing frequency of PT
- Nonfatal SAEs by decreasing frequency of PT
- Exposure-adjusted SAEs
- TEAE leading to drug discontinuation by SOC and PT
- TEAE leading to drug discontinuation of Lenvatinib by SOC and PT
- TEAE leading to drug discontinuation of Everolimus/Pembrolizumab by SOC and PT
- TEAE leading to drug reduction and/or interruptions by SOC and PT
- TEAE leading to drug reduction and/or interruptions of Lenvatinib by SOC and PT
- TEAE leading to drug reduction and/or interruptions Everolimus/Pembrolizumab by SOC and PT

The following subject AE listings (treatment-emergent or otherwise) will be provided:

- All AEs with fatal outcome
- All SAEs
- All AE leading to study drug discontinuation
- All AEs leading to study drug dose reduction and/or interruption

5.6.2.2 Treatment-Emergent Adverse Events of Clinical Interest

Clinically significant TEAEs (CSAE) for Lenvatinib, TEAEs of special interest (AEOSI) for everolimus, pembrolizumab, respectively, will be identified based on the review of safety data by Clinical and Pharmacovigilance.

CSAE and AEOSI will be summarized by the grouped terms for any Grade and Grade ≥3.Time to onset of the CSAE and AEOSI will be summarized using descriptive statistics for subjects with CSAE and AEOSI reported, median time to first onset, proportion of subjects with CSAE and AEOSI leading to drug withdrawal or dose modifications will be summarized for any Grade and Grade ≥ 3 CSAE and AEOSI. Time to onset of the CSAE and AEOSI is defined as the time from first dose to the first onset of the CSAE and AEOSI.

5.6.3 Laboratory Values

Laboratory results will be summarized using Système International (SI) units, as appropriate. Laboratory values that are non-missing and reported as 'below the detectable limit' of an assay will be replaced by half the detectable limit in the summary tables. On treatment laboratory tests will be defined as laboratory tests conducted from the start of treatment to no more than 30 days after the last dose of study treatment. Central laboratory test results will be primary data source for laboratory analyses. Only when the central laboratory tests results are missing, the local laboratory test results will be used as substitute. Other laboratory parameters collected for some individual subjects will be presented in listing only. Box and Whisker plots will also be generated for individual laboratory parameters and change from baseline by visit for all visits with at least 10% of subjects in all treatment arms.

For all quantitative parameters listed in protocol Section 9.5.1.4 Safety Assessments (Laboratory Measurements), the actual value and the change from baseline to each postbaseline visit and to the end of treatment (defined as the last on-treatment value) will be summarized by visit and treatment group using descriptive statistics. Qualitative parameters listed in protocol Section 9.5.1.4 will be summarized using frequencies (number and percentage of subjects), and changes from baseline to each postbaseline visit and to end of treatment will be reported using shift tables. Percentages will be based on the number of subjects with both nonmissing baseline and relevant postbaseline results.

5.6.3.1 Hematology Clinical Chemistry

Laboratory parameters that are graded in CTCAE version 4.03 will be summarized by CTCAE grade. In the summary of laboratory parameters by CTCAE grade, for parameters with CTCAE grading in both high and low direction (*e.g.*, calcium, glucose, magnesium, potassium, sodium), CTCAE in high and low directions will be summarized separately. Subjects with treatment-emergent markedly abnormal laboratory values (TEMAV) will also be summarized. Markedly abnormal is defined as a value that is above or below the normal range and the CTCAE grade increased from baseline by 2 or more grades, except for phosphate which must have shifted by 3 or more grades.

Laboratory test results will be reported in the following methods with three categories: Hematology, liver and renal, and other clinical chemistry.

- Descriptive summary statistics for all parameters and their changes from baseline will be summarized using number of subjects, mean, SD, median, Q1, Q3, and range by visit.
- Shifts from baseline to the worst postbaseline CTC grade will also be presented.

A listing of potential Hy's law cases who meets one of the following criteria: (aspartate aminotransferase or alanine transferase)>3xULN and total bilirubin ≥2xULN and alkaline phosphatase <2xULN at any postbaseline assessment time point during the study will be provided. Parameters selected per Hy's Law definition were provided in FDA Guidance for Industry Drug Induced Liver Injury: Premarketing Clinical Evaluations, July 2009.

5.6.3.2 Urinalysis

Proteinuria: Shifts from Baseline to worst postbaseline for proteinuria determined by dipstick (negative, trace, 1+, 2+, 3+, 4+) will be presented by treatment arms. The urinalysis results from UPCR or 24-hour urine protein (< 1.0, 1.0–3.5, > 3.5 g/24 hrs according to CTCAE grade) will also be summarized (frequency and percentage) by worst Grade.

5.6.4 Vital Signs

Descriptive statistics for vital signs parameters (diastolic and systolic blood pressure, pulse, respiration rate, temperature, weight) and changes from baseline will be presented by visit and treatment group.

Blood pressure will also be summarized using a shift table from baseline to worst postbaseline by categories defined based on CTCAE grades (Table 5).

Table 5 Blood Pressure CTCAE Grades

Grade	Blood Pressure (mm Hg)	
	Systolic	Diastolic
0 (Normal)	≤119	≤79
1 (Prehypertension)	120 – 139	80 – 89
2 (Stage 1 Hypertension)	140 - 159	90 – 99
3 (Stage 2 Hypertension)	≥ 160	≥ 100

For each subject at each visit, SBP and DBP will be graded separately. Then the blood pressure grade will be the worst of the two grades: SBP grade and the DBP grade. The worst blood pressure category per subject during treatment will be used in the shift table.

Subjects with SBP ≥ 140 mm Hg or DBP ≥ 90 mm Hg on treatment will be listed along with their hypertensive medications.

5.6.5 Electrocardiograms

ECG assessments were performed at Day 1 of each cycle. Descriptive statistics for ECG parameters (QTcB [ms] and QTcF [ms]) and changes from baseline will be presented by visit and treatment group.

Shift tables will present changes from baseline in ECG interpretation (categorized as normal; abnormal, not clinically significant; and abnormal, clinically significant) to end of treatment (or end of phase or by visit).

In addition, the number (percentage) of subjects with at least 1 postbaseline abnormal ECG result in QTc Fridericia during the treatment period will be summarized. Clinically abnormal ECG results in QTc Fridericia will be categorized as follows:

Absolute QTc interval prolongation:

QTc interval >450 ms

QTc interval >480 ms

QTc interval >500 ms

Change from baseline in QTc interval:

QTC interval increases from baseline >30 ms

QTC interval increases from baseline >60 ms

5.6.6 Other Safety Analyses

Karnofsky Performance Status (KPS) will be summarized by a shift table from baseline to worst postbaseline scale and by frequency counts for each category for each visit.

LVEF (%) obtained by echocardiography will be summarized. The lowest postbaseline value and change from baseline will be summarized with n, mean, SD, median, Q1, Q3, and range. A subject data listing will be provided.

In addition, LVEF quantitative values will be classified into the following 5 categories:

Hyperdynamic = LVEF greater than 70%

Normal = LVEF 50% to 70% (midpoint 60%)

Mild dysfunction = LVEF 40% to 49% (midpoint 45%)

Moderate dysfunction = LVEF 30% to 39% (midpoint 35%)

Severe dysfunction = LVEF less than 30%

A shift table from baseline to the worst postbaseline values will be generated.

5.7 Other Analyses

A detailed analysis plan for health-related quality of life will be provided in a separate document. Protocol deviations and adverse events associated with COVID-19 (coronavirus disease 2019) will be presented.

5.8 Exploratory Analyses

Exploratory analyses may be conducted as appropriate. Any exploratory analyses that are performed will be appropriately titled and labeled as exploratory and will be clearly distinguished from planned analyses when results are reported in the clinical study report.

5.9 Extension Phase Analyses

A separate SAP will be written for Extension Phase analyses, if needed.

6 INTERIM ANALYSES

Interim analyses of PFS, OS, and ORR are planned in this study. The interim analyses will be conducted by an independent statistical group (eg, a CRO) that has no other responsibilities for the study. The timing of each analysis are summarized in Table 6. Type I error control for the efficacy analyses as well as efficacy boundaries are described in Section 5.3.3.

The safety monitoring and interim analyses for PFS and OS will be conducted by the independent Data Monitoring Committee (DMC). The recommendation as to whether to stop the trial will be reached by the DMC based on their review of safety data and efficacy data with treatment information. The function and membership of the DMC will be described in the DMC Charter. The external DMC will conduct regular safety monitoring. The timing of the safety monitoring will be specified in the DMC Charter.

Table 6 Summary of Interim and Final Efficacy Analyses

No.	Analysis	Endpoint(s)	Timing	Estimated Time after First Subject Randomized
1	Interim analysis of ORR and DOR (the first 88 subjects from Arm B)	ORR DOR	Median follow-up of 12 months and a minimum DOR follow-up of 6 months	~28 months
2	Interim analysis of PFS, Interim analysis of OS	PFS OS ORR*	Trigger: approximately 4 months after the last subject randomized and approximately 310 (80% IF) PFS events observed in Arms B and C (estimated to have ~140 (45% IF) deaths observed for each comparison)	~38 months
3	Final analysis of PFS, Interim analysis of OS	PFS OS	Trigger: ~ 388 PFS events observed for each comparison (estimated to have 182 (60% IF) deaths observed for each comparison)	~45 months
4	Interim analysis of OS	OS	Trigger: ~243 (80% IF) deaths observed for each comparison	~57 months
5	Final analysis of OS	OS	Trigger: ~304 deaths observed for each comparison	~69 months

DOR = duration of response; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; IF=information fraction.

*: The p-value for hypothesis testing of ORR will be based on the ORR data at the analysis No 2.

7 CHANGES IN THE PLANNED ANALYSES

Not applied.

8 DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

8.1 EFFICACY DATA HANDLING

8.1.1 Visit Windows for Efficacy Analysis

Tumor assessments were to be performed every eight weeks during the Randomization Phase and according to the local standard of care in the Extension Phase (or sooner if clinically indicated) until documentation of disease progression. Time window for radiologic assessment is $\pm$5 days.

For definitions, derived variables and data sets, please refer to the data set specifications for more details. The following points should be noted:

- Confirmed CR/PR and unconfirmed CR/PR will be derived.

- Early death will refer to the death which occurred within 121 days after randomization.

  Date of death – Date of randomization ≤ 121 days

8.1.2 Missing/Partial Dates

Refer to programming specifications for details to handle any missing or partial dates.

8.2 Safety and Disease History Data Handling

8.2.1 Visit Windows for Safety Analyses

Visit windows will be defined to be upper and lower bounds of one day of the scheduled visit, following the protocol, which states that efforts should be made to conduct laboratory assessments and administer treatment on the day scheduled (± 1 day). In the calculation of descriptive statistics for laboratory values and vital signs, if a visit has multiple observations, the observation closest in date and time to the target visit day will be used in the analysis. If two or more observations have the same distance to the target visit day, the one that has the highest CTCAE grade or is furthest away from the normal range will be used.

The purpose of this windowing is to provide a single record per subject per visit for the calculation of descriptive statistics per scheduled visit, and change from baseline per visit. Other safety analyses (e.g., worst grade laboratory results) will include all observations.

8.2.2 Definitions, Derived Variables, and Data Sets

Baseline is defined as the non-missing value most recently collected before the first dose.

The following factors will convert days to months or years:

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

Time from first diagnosis to randomization (years) is:

(Date of randomization – Date of first diagnosis) / 365.25.

Time from first metastatic diagnosis to randomization (years) is:

(Date of randomization – Date of first metastatic diagnosis) / 365.25.

The Cockcroft and Gault formula (Cockcroft & Gault, 1976) for creatinine clearance (mL/min) calculation is:

Males:

$$\frac{(140 - \text{age (years)}) \times \text{weight (kg)}}{\text{Serum creatinine (mg/dL)} \times 72},$$

Females:

$$\frac{(140 - \text{age (years)}) \times \text{weight (kg)}}{\text{Serum creatinine (mg/dL)} \times 72} \times 0.85.$$

8.3 Pharmacokinetics/Pharmacodynamics Data Handling

Details on calculating pharmacokinetics parameters and the way that below limit of quantification (BLQ) values will be replaced will be detailed in a separate plan detailing the analyses to be conducted of the exploratory pharmacokinetics/pharmacodynamics exposure-response relationships.
When developing individual concentration-time profiles, BLQ values will be replaced with zero for the linear plot or missing for the semi-logarithm plot, respectively.

When calculating the mean or median value for the concentration at a given time point, the BLQ values will be assigned as zero. If the proportion of values reported as BLQ is more than 50%, no summary statistics should be represented at that time point, and the value will be treated as missing in mean or median concentration profiles.

9 PROGRAMMING SPECIFICATIONS

The rules for programming derivations and dataset specifications are provided in separate documents.

10 STATISTICAL SOFTWARE

Statistical programming and analyses will be performed using SAS® (SAS Institute, Inc., Cary, NC, USA), version 9.1 or higher, and/or other validated statistical software as required.

11 MOCK TABLES, LISTINGS, AND GRAPHS

The study TLG shells will be provided in a separate document, which will show the content and format of all tables, listings, and graphs in detail.

12 REFERENCES

Cockcroft D, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron. 1976;16(1): 31-41.

Common Terminology Criteria for Adverse Events (CTCAE). Version 4.03. United States Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, USA, May 28, 2009.

Eisenhauer EA, Therasse P, Bogaerts J Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45: 228-47.

EMA Guideline on the Evaluation of Anticancer Medicinal Products in Man, July 2013. Available from: https://www.ema.europa.eu/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man_en.pdf

FDA Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, May 2007 Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf.

FDA Guidance for Industry Drug Induced Liver Injury: Premarketing Clinical Evaluations, July 2009. Available from: https://www.fda.gov/downloads/guidances/UCM174090.pdf

Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions, 3rd ed. New York: Wiley, 2003.

Ko JJ, Xie W, Kroeger N, Lee JL, Rini BI, Knox JJ, et al. The International Metastatic Renal Cell Carcinoma Database Consortium model as a prognostic tool in patients with metastatic renal cell carcinoma previously treated with first-line targeted therapy: a population-based study. Lancet Oncol 2015; 16: 293–300.

Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958; 53: 457-481.

Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research, 2013; 5: 311-320.

13 APPENDICES

13.1 Heng Score Criteria

International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) Model Criteria Determination of Prognostic Score in First Line Setting (Ko et al, 2015)

Parameter	Criteria Value	Subject Value	If subject value meets criteria value, enter 1
KPS, %	< 80		
Time from diagnosis to treatment with systemic therapy, months	< 12		
Hemoglobin	< LLN		
*Corrected Calcium, mg/dL	> ULN		
Neutrophil count	> ULN		
Platelet count	> ULN		
			Sum total of above= IMDC Prognostic Score

Corrected Calcium (mmol/L) = Ca measured (mmol/L) + 0.025 (40 – albumin (g/L))

*The formula is not applicable when serum albumin concentration is normal (>40 g/L); in such situations, the total (uncorrected) serum calcium should be used instead.

Risk Group Based on IMDC model

Risk Group	IMDC Prognostic Score
Favorable-Risk	0
Intermediate-Risk	1 or 2
Poor-Risk	3-6

SIGNATURE PAGE

Author:

Date

Oncology Business Group, Eisai Inc.

Approval:

Date

Oncology Business Group, Eisai Inc.

Date

Oncology Business Group, Eisai Ltd.