
# STATISTICAL ANALYSIS PLAN

| | |
|---|---|
| **Study Protocol Number:** | E7080-G000-218 |
| **Study Protocol Title:** | A Randomized, Open-Label (formerly Double-Blind), Phase 2 Trial to Assess Safety and Efficacy of Lenvatinib at Two Different Starting Doses (18 mg vs 14 mg QD) in Combination With Everolimus (5 mg QD) in Renal Cell Carcinoma Following One Prior VEGF-Targeted Treatment |
| **Date:** | 29 Apr 2020 |
| **Version:** | 2.0 |

# REVISION HISTORY

Revisions per Version 2.0
Date: 29 Apr 2020

| Change | Rational | Affected Sections |
|---|---|---|
| Updated that the data cutoff for the primary analysis refers to the end of the randomization phase) and that the End of Study refers to the last subject last visit after which all subjects will have completed their off-treatment visits.<br><br>The term "final" analysis was changed to "primary" analysis according to the Protocol Amendment 07. | Protocol was amended (Protocol Amendment 07), SAP updates were made accordingly. | Section 3.2 |
| For the stratification factor of MSKCC prognostic group from IxRS, some subjects with MSKCC Prognostic Score 2 were misclassified to intermediate risk on Interactive Response Technology (IRT) system which should have been classified to poor risk per protocol. Poor risk will be pooled with the intermediate risk in the relevant stratified analyses based on IxRS data, the favorable risk will remain intact. | For the stratification factor of MSKCC prognostic group from IxRS, some subjects with MSKCC Prognostic Score 2 were misclassified to intermediate risk on IRT system which should have been classified to poor risk per protocol. It is then determined that poor risk group will be pooled with the intermediate risk group in the relevant stratified analyses based on IxRS data, the favorable risk group will remain intact. | Section 5.3.2; Section 5.4 |
| Added hazard ratio estimates for PFS, OS and PFS2 analyses. | Added hazard ratio estimates for PFS, OS and PFS2 analysis for descriptive purpose. | Section 5.4.2.1, 5.4.2.3, and 5.4.2.4 |
| Changed the major protocol deviations to important protocol deviations. | Per new Standard Working Practice (SWP) 101-206.00-SWP, protocol deviation type will be classified as important or minor. | Section 5.2.1, 5.2.3 |

# 1 TABLE OF CONTENTS

# 2 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

| Abbreviation | Term |
|---|---|
| AE | adverse event |
| AEoSI | AEs of special interest |
| AJCC | American Joint Committee on Cancer |
| ATC | anatomical therapeutic class |
| BLQ | below limit of quantification |
| BOR | best overall response |
| BMI | body mass index |
| CBR | clinical benefit rate |
| CI | confidence interval |
| CMH | Cochran-Mantel-Haenszel |
| CMQ | customized MedDRA queries |
| CR | complete response |
| CRF | case report form |
| CSAEs | clinically significant adverse events |
| CSR | clinical study report |
| CV | coefficient of variation |
| DCR | disease control rate |
| DMC | data monitoring committee |
| DOR | duration of response |
| ECGs | electrocardiograms |
| EMA | European Medicines Agency |
| 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 Symptom Index-Disease Related Symptoms |
| HRQoL | Health-Related Quality of Life |
| ICL | imaging core laboratory |
| IIR | independent imaging review |

| Abbreviation | Term |
|---|---|
| IRT | interactive response technology |
| IxRS | interactive voice and web response system |
| LDH | lactate dehydrogenase |
| LLT | lower level term |
| LVEF | left ventricular ejection fraction |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MUGA | multiple gate acquisition |
| NA | not applicable |
| NE | not evaluable |
| NYHA | New York Heart Association |
| KPS | Karnofsky performance status |
| MSKCC | Memorial Sloan-Kettering Cancer Center |
| OR | odds ratio |
| ORR | objective response rate |
| $ORR_{24w}$ | objective response rate as of Week 24 |
| OS | overall survival |
| PFS | progression-free survival |
| PFS2 | progression-free survival on next line of therapy |
| PD | progressive disease or pharmacodynamic |
| PK | pharmacokinetic |
| PO | orally |
| PR | partial response |
| PT | preferred term |
| QD | once daily |
| QTc | corrected QTc interval |
| QTcF | corrected for QTc interval using Frederica's correction factors |
| RECIST | Response Evaluation Criteria in Solid Tumors |
| RBC | red blood cell |
| RCC | renal cell carcinoma |
| SAE | serious adverse event |

| Abbreviation | Term |
|---|---|
| SAP | statistical analysis plan |
| SD | standard deviation or stable disease |
| SE | standard error |
| SGQs | sponsor-generated queries |
| SI | Système International |
| SMQs | standardized MedDRA queries |
| SOC | system organ class |
| SWP | standard working practice |
| TEAE | treatment-emergent adverse event |
| TEMAV | treatment-emergent markedly abnormal values |
| TLGs | tables, listings, and graphs |
| TNM | tumor-node-metastasis |
| TSH | thyroid stimulating hormone |
| VEGF | vascular endothelial growth factor |
| WBC | white blood cell |
| 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 of the primary analysis for Eisai Protocol E7080-G000-218 (Amendment 07). Analyses on subjects' Health-Related Quality of Life (HRQoL) data, pharmacokinetic (PK)/pharmacodynamics (PD), biomarkers, and relationships between PK and efficacy and safety will be included in separate statistical analysis plans.

## 3.1 Study Objectives

### 3.1.1 Primary Objective

The primary objective of this study is to assess whether a starting dose of lenvatinib 14 mg in combination with everolimus 5 mg once daily (QD) will provide comparable efficacy (based on objective response rate [ORR] at 24 weeks [$ORR_{24W}$]) with an improved safety profile compared to lenvatinib 18 mg in combination with everolimus 5 mg (based on treatment-emergent intolerable Grade 2, or any ≥ Grade 3 adverse events (AEs) in the first 24 weeks after randomization).

### 3.1.2 Secondary Objectives

The secondary objectives of the study are:

- To assess progression-free survival (PFS)
- To assess ORR
- To determine the tolerability and safety profile of lenvatinib in combination with everolimus
- To assess the proportion of subjects who discontinued treatment due to toxicity
- To assess time to treatment failure due to toxicity
- To assess pharmacokinetic (PK) profiles of lenvatinib and everolimus during combination therapy and to assess PK and pharmacodynamics (PD) drug-drug interactions
- To evaluate overall survival (OS)
- To evaluate the impact of disease and treatment on subjects' Health-Related Quality of Life (HRQoL) as assessed by using the Functional Assessment of Cancer Therapy - Kidney Symptom 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
- To evaluate the PFS on next line of therapy (PFS2)

### 3.1.3 Exploratory Objectives

The exploratory objectives of the study are:

- To explore tumor response parameters ($ORR_{24W}$, ORR, PFS) based on blinded independent imaging review (IIR) for efficacy assessment
- To explore blood biomarkers that correlate with efficacy-related endpoints of this study
- To develop exposure/biomarker/clinical endpoint models (whenever possible, using a mechanism-based approach) for both efficacy and safety data that will allow exploration of alternative dosing regimens with a better efficacy/safety profile than the lenvatinib 18 mg plus everolimus 5 mg dose

## 3.2 Overall Study Design and Plan

Study E7080-G000-218 was designed as a multicenter, randomized, open-label (formerly double-blind) study, conducted as a postmarketing requirement of the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) to evaluate an alternate dose regimen for lenvatinib in combination with everolimus. Lenvatinib 18 mg daily in combination with everolimus 5 mg daily is approved in the US and EU for the treatment of adult subjects with advanced renal cell carcinoma (RCC) following 1 prior vascular endothelial growth factor (VEGF) targeted therapy.

This study will evaluate the combination of lenvatinib and everolimus at a 14 mg starting dose of lenvatinib and allow up-titration of lenvatinib to determine whether this alternate dose regimen provides comparable efficacy but has a better safety profile than the 18 mg starting dose in this subject population. The 14 mg starting dose will be escalated to 18 mg if no Grade 2 (intolerable) or any ≥ Grade 3 treatment-emergent adverse events (TEAEs) that require dose reduction are observed in the first cycle (4 weeks) of treatment. If Grade 2 (intolerable) or Grade 3 or 4 TEAEs are observed, the lenvatinib dose will be reduced, as described in the dose reduction section in protocol Section 9.4.2.1. Both lenvatinib and everolimus will be administered orally (PO) and QD.

Eligible subjects will have measurable disease according to Response Evaluation Criteria in Solid Tumors Version 1.1 (RECIST 1.1) and will be randomly assigned to each treatment arm in a 1:1 ratio. The total sample size will be approximately 338 subjects. Randomization will follow a predefined randomization scheme based on the following stratification factors: Memorial Sloan-Kettering Cancer Center (MSKCC) prognostic groups (favorable, intermediate, or poor risk); and whether subjects have had a prior PD-1/PD-L1 treatment (yes or no).

Subjects will receive study treatment as continuous 28-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 in Table 6 in protocol. Subjects will discontinue study treatment upon evidence of progressive disease, as judged by the investigator. After disease progression, subjects will be followed for survival and PFS2.

This study consists of 2 phases, the Pre-randomization Phase and the Randomization Phase, as shown in Figure 1:

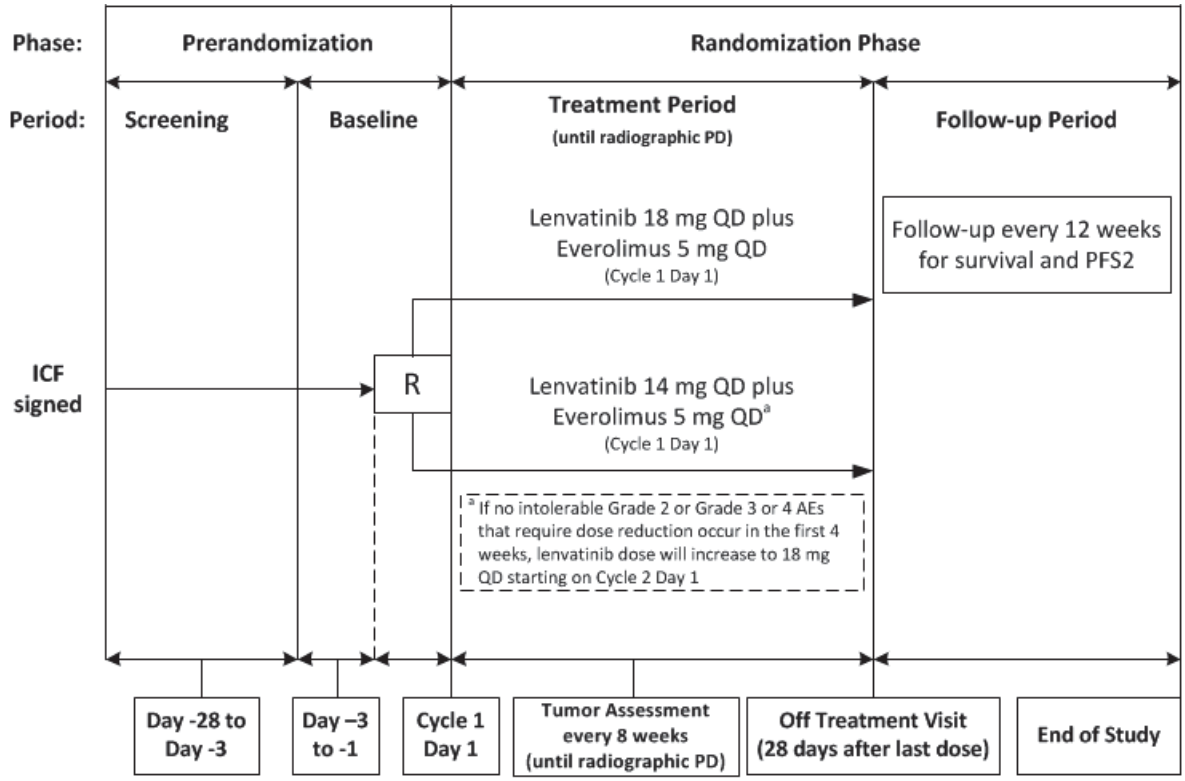


**Figure 1 Study Design for Study E7080-G000-218**
R = randomization
PD = progressive disease
PFS2 = PFS on next line of therapy

- The Pre-randomization Phase will last no longer than 28 days and will include a Screening Period to establish protocol eligibility and a Baseline Period to confirm eligibility and establish disease characteristics prior to randomization and treatment.

- The Randomization Phase will consist of a Treatment Period and a Follow-up Period. It will begin at the time of randomization of the first subject and will end at the data cutoff for the primary analysis, which is defined as when all randomized subjects complete the Week-24 tumor assessments or discontinue study treatment before Week 24.

- The Treatment Period will begin with the first dose of study drug administration in Cycle 1 and continue in 28-day cycles until completion of the off-treatment assessments (within 28 days after the last study drug administration). Subjects will undergo safety and efficacy assessments as defined in the Schedule of Procedures/Assessments in protocol Table 6. Subjects will continue to receive study treatment until confirmed disease progression, development of unacceptable toxicity, subject request, withdrawal of consent, or study termination by the sponsor. Subjects who discontinue treatment before the data cut off for the primary analysis will enter the Follow-up Period, and will be followed every 12 weeks (±1 week) after the off-treatment visit. If a clinic visit is not feasible, follow-up information may be obtained via telephone or email.

- The Follow-up Period will begin immediately after the off-treatment assessments have been completed and will continue as long as the study subject is alive, until the study subject withdraws consent, or until the data cutoff for the primary analysis. Subjects who discontinue study treatment before disease progression will continue to undergo tumor assessments every 8 weeks and as of Amendment 06, send these to the imaging core laboratory (ICL) until documentation of disease progression or start of another anticancer therapy. Following the off-treatment visit, subjects will continue to be followed every 12 weeks (±1 week) for survival and PFS2, and all anticancer treatments received will be recorded until the data cutoff for the primary analysis. This information will be recorded unless the information is not allowed to be provided due to confidentiality.

The data cutoff for the primary analysis will occur at the end of the Randomization Phase, which is defined as the time when all randomized subjects have completed Week 24 assessments or have discontinued study treatment prior to Week 24.

Subjects will continue to receive investigational product until they complete the off-treatment visit prior to their transition to commercial lenvatinib and everolimus or an access program. The last subject last visit for the End of Study will be the date of the off-treatment visit for the last subject.

# 4 DETERMINATION OF SAMPLE SIZE

The objective of the study is to assess whether a starting dose of lenvatinib 14 mg QD in combination with everolimus 5 mg will provide comparable efficacy with an improved safety profile compared to lenvatinib 18 mg in combination with everolimus 5 mg. Determination of whether 14 mg lenvatinib can be used as an alternative dosing strategy will be based on clinical judgment by the Sponsor in consultation with the independent Data Monitoring Committee (DMC) as specified in DMC Charter by assessing risks and benefits according to the totality of data at either of the interim or final analyses. Nevertheless, the sample size is guided by the

plan of testing non-inferiority on the primary efficacy endpoint and superiority on the primary safety endpoint. Superiority on the primary safety endpoint will only be tested if the non-inferiority on the primary efficacy endpoint is claimed. The details and assumptions are provided below.

Sample size is based on detecting both the non-inferiority of $ORR_{24W}$ and superiority of the primary safety endpoint of the proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization in comparison of the lenvatinib 14 mg arm to the lenvatinib 18 mg arm.

The "final analysis" stated in the sections of Determination of Sample Size, Primary Efficacy Analyses, Interim Analyses, and Appendix Section 13.1 refers to the last analysis when the statistical inference is made for the study, that is, at the data cutoff for the primary analysis.

#### NON-INFERIORITY OF $ORR_{24W}$ COMPARING LENVATINIB 14 MG TO LENVATINIB 18 MG

Based on the assumption from Study E7080-G000-205 that the confirmed ORR for the 18 mg lenvatinib + 5 mg everolimus arm is 37% (19 responders out of N=51) vs 6% for the everolimus arm (3 responders out of N=50), the 95% confidence interval (CI) of the odds ratio (OR) comparing everolimus vs. lenvatinib 18 mg + everolimus arm is (0.029, 0.395). The non-inferiority margin is chosen to ensure that a reasonable fraction of the lenvatinib 18 mg + everolimus vs. everolimus treatment effect is preserved. A 70% retention of the treatment effect of lenvatinib 18 mg + everolimus vs. everolimus is used for this design. Following the approach in Rothmann, et al. (2003), using the 95% CI upper limit method based on logarithm of the odds ratio, the non-inferiority margin is estimated as:

$$\exp((1-\delta) * (\text{upper limit of 95\% CI of log OR (everolimus/lenvatinib 18 mg + everolimus)}))$$

where $0<\delta< 1$ is the retention rate.

To retain 70% of the lenvatinib 18 mg + everolimus vs. everolimus treatment effect, the non-inferiority margin (M) of the odds ratio is estimated to be:

$M = \exp((1 - 0.7) * \log(0.395)) = 0.76$ (i.e, $H_a$: OR (14 mg/18 mg)>M).

Table 1 below lists the non-inferiority margins on the scale of difference in $ORR_{24W}$ between lenvatinib 14 mg + everolimus arm and lenvatinib 18 mg + everolimus arm corresponding to a 0.76 non-inferiority margin on the odds ratio scale for a different $ORR_{24W}$ in lenvatinib 18 mg + everolimus arm.

**Table 1 Non-inferiority Margins on Scale of Treatment Difference in $ORR_{24W}$**

| $ORR_{24W}$ in lenvatinib 18 mg arm | Non-inferiority Margin on Difference Scale (lenvatinib 14 mg arm - lenvatinib 18 mg arm) |
|---|---|

| | |
|---|---|
| 10% | -2% |
| 20% | -4% |
| 30% | -5% |
| 40% | -6% |
| 50% | -7% |
| 60% | -7% |

Two interim analyses will take place when 150 and 200 total subjects in the Per-Protocol Analysis Set 1 have completed 24 weeks follow-up or discontinue earlier. Each interim analysis will test both non-inferiority and futility of the 14 mg arm $ORR_{24W}$ compared to the 18 mg arm $ORR_{24W}$. An O'Brien-Fleming stopping boundary will be used for non-inferiority. An interpolated non-binding stopping boundary will be used for futility, which will spend $\beta=0.005$ and $\beta=0.10$ at the first and second interim analysis, respectively. Assuming 37% $ORR_{24W}$ in the lenvatinib 18 mg arm and 45% $ORR_{24W}$ in the lenvatinib 14 mg arm, and adjusting for the interim analyses, a total of 306 subjects (153 per arm) in the Per-Protocol Analysis Set 1 is required to achieve 80% statistical power at 1-sided $\alpha=0.05$.

The stopping boundaries on the *P* value scale and the cumulative error probabilities spent at each interim analysis and final analysis are shown in Table 2 below. For example, at the second interim analysis, non-inferiority in $ORR_{24W}$ will be claimed if the 1-sided *P* value is ≤0.014; futility will be claimed if the 1-sided *P* value is ≥0.207.

**Table 2 Stopping Boundaries**

| Analysis # | Cumulative α Spent | Efficacy Boundary (*P* value) | Cumulative β Spent | Futility Boundary (*P* value) |
|---|---|---|---|---|
| Interim Analysis #1 | 0.005 | 0.005 | 0.005 | 0.776 |
| Interim Analysis #2 | 0.015 | 0.014 | 0.10 | 0.207 |
| Final Analysis | 0.05 | 0.045 | 0.2 | 0.045 |

#### SUPERIORITY OF PRIMARY SAFETY ENDPOINT COMPARING LENVATINIB 14 MG TO LENVATINIB 18 MG

At each interim analysis and the final analysis in the Per-Protocol Safety Analysis Set, superiority on the primary safety endpoint will be evaluated if the non-inferiority boundary is crossed. Assuming 75% subjects with intolerable Grade 2 or above TEAEs within 24 weeks after randomization in the 18 mg arm, with a total of 306 subjects, a superiority test at 2-sided $\alpha=0.05$ will give 80% statistical power to detect a 15% drop in proportion of subjects with intolerable Grade 2 or above TEAEs within 24 weeks after randomization in the 14 mg-arm.

In consideration of both the primary efficacy and safety endpoints, a total of approximately 306 subjects were originally planned to be randomized in a 1:1 ratio to both treatment arms. Since there were 32 subjects who received ≥2 incorrect lenvatinib doses due to IxRS issues, the number of subjects to be randomized will be increased by 32 to a total of approximately

338. Therefore, there will be approximately 306 subjects in the Per-Protocol Analysis Set 1. Randomization will be stratified by MSKCC prognostic group (favorable, intermediate, and poor risk) and whether subjects have had a prior PD-1/PD-L1 treatment (yes or no).

# 5 STATISTICAL METHODS

In general, continuous variables will be summarized using descriptive statistics such as mean, standard deviation (SD), median, and range (minimum and maximum). Categorical variables will be summarized using frequency and percentage. For time-to-event variables, which is defined as the time from the date of randomization to the date of the event, the Kaplan-Meier method will be used for descriptive summaries. For the calculation of time-to-event or duration-of-event variables for on-study events, the difference between the start date and the end date plus 1 day will be used. For durations of events (eg, baseline disease characteristics or prior therapies) prior to randomization, the durations will be calculated as the date of randomization minus the date of the event, the details are specified in Section 8.4.1.

## 5.1 Study Endpoints

### 5.1.1 Primary Endpoints

The primary endpoints are:

- $ORR_{24W}$ as assessed by investigator according to RECIST 1.1. $ORR_{24W}$ is defined as the proportion of subjects with best overall response (BOR) of complete response (CR) or partial response (PR) as of the Week 24 (after randomization) timepoint during treatment or within 28 days after the last dose date but on or prior to the start of new anticancer therapy. To be considered as BOR, all responses must be confirmed no less than 4 weeks after the initial assessment of response.

- Proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization (as of the Week 24 timepoint).

### 5.1.2 Secondary Endpoints

The secondary efficacy endpoints are defined as follows:

- PFS, defined as the time from the date of randomization to the date of first documentation of disease progression or date of death, whichever occurs first.

- ORR as assessed by investigator according to RECIST 1.1, defined as the proportion of subjects with BOR of CR or PR during treament or within 28 days after the last dose date but on or prior to the start of new anticancer therapy. To be considered as BOR, all responses must be confirmed no less than 4 weeks after the initial assessment of response.

- Overall safety profile and tolerability of lenvatinib in combination with everolimus.
- Proportion of subjects who discontinue treatment due to toxicity, defined as the proportion of subjects who discontinue study treatment due to TEAEs.
- Time to treatment failure due to toxicity, defined as the time from the date of randomization to the date that a subject discontinues study treatment due to TEAEs.
- Lenvatinib and everolimus exposure parameters and PK and PD drug-drug interactions.
- OS, measured from the date of randomization until date of death. In the absence of confirmation of death, subjects will be censored either at the date that the subject was last known to be alive or the date of data cutoff, whichever comes earlier.
- HRQoL will be assessed using the FKSI-DRS, the EORTC QLQ-C30 and the EuroQol EQ-5D-3L instruments.
- PFS2, defined as the time from randomization to the date of disease progression on next line of therapy or death, whichever occurs first.

### 5.1.3 Exploratory Endpoints

The exploratory endpoints include:

- Tumor response endpoints $ORR_{24W}$, ORR, and PFS based on IIR assessments. These endpoints will be defined in the same way as those based on investigator assessments.
- Associations between blood biomarker and efficacy related endpoints.
- Development of exposure/biomarker/clinical endpoint models (whenever possible, using a mechanism-based approach) for both efficacy and safety data.

## 5.2 Study Subjects

### 5.2.1 Definitions of Analysis Sets

- Full Analysis Set will include all randomized subjects. This will be a secondary analysis set for efficacy endpoints, which will be analyzed according to the treatment arm to which subjects are randomized, regardless of the treatment actually received.
- Per-Protocol Analysis Set 1 will include all randomized subjects minus the 32 subjects who had received ≥2 incorrect lenvatinib doses due to IxRS issues. This will be the primary analysis set for efficacy endpoints, which will be analyzed according to the treatment arm to which subjects are randomized.

- Per-Protocol Analysis Set 2 will include all subjects who received at least 1 dose of study drug, had no important protocol deviations, and had both baseline and at least 1 postbaseline tumor assessment. Subjects who died before the first postbaseline tumor assessment will also be included. The Per-Protocol Analysis Set 2 will be a secondary analysis set for efficacy endpoints. The 32 subjects who received ≥2 incorrect lenvatinib doses due to IxRS issues are considered as having experienced important protocol deviations and will be excluded from the Per-Protocol Analysis Set 2. (A 33rd subject received a single incorrect lenvatinib dose due to IxRS issues, but given the brief exposure of the incorrect [slightly higher] dose before detection and correction, and because there were no adverse effects, this subject will be considered to have experienced a minor protocol deviation and will not be excluded from analysis sets.)

- Safety Analysis Set will include all subjects who were randomized and received at least 1 dose of study drug. This will be the analysis set for all safety evaluations, which will be analyzed according to the treatment actually received.

- Per-Protocol Safety Analysis Set will include all treated subjects in Per-Protocol Analysis Set 1. This will be the primary analysis set for the primary safety endpoint, which will be analyzed according to the treatment actually received.

- Pharmacokinetic Analysis Set will include all subjects who received at least 1 dose of study drug with documented dosing history and had at least 1 evaluable lenvatinib plasma or everolimus whole blood concentration data.

- Pharmacodynamic Analysis Set will include all subjects who received at least 1 dose of study drug with documented dosing history and had at least 1 evaluable sample for pharmacodynamics data.

  For the Pharmacokinetic, Pharmacodynamic and other Biomarker endpoints, their respective analysis plans will specify if the analysis set will or will not include the 32 subjects who received ≥2 incorrect lenvatinib doses due to IxRS issues.

- Quality of Life (QoL) Analysis Set will consist of all subjects who had any QoL data.

**Table 3 Analysis Sets in Data Analyses**

| Tables | All Screened Subjects | Full | Per-Protocol Analysis Set 1 | Per-Protocol Analysis Set 2 | Safety | Per-Protocol Safety Analysis Set | PK | PD | QoL |
|---|---|---|---|---|---|---|---|---|---|
| Protocol Deviations | | • | | | | | | | |
| Disposition | • | • | • | • | | | | | |
| Demography & Baseline Characteristics | | • | • | • | | | | | |
| Disease History | | • | | | | | | | |
| Prior and concomitant medication | | • | • | | | | | | |
| New anticancer therapy | | • | • | | | | | | |
| Efficacy analysis | | • | • | • | | | | | |
| Primary safety endpoint - proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization | | | | • | • | • | | | |
| Safety analysis (including drug exposure, AEs, laboratory tests, vital signs, etc.) | | | | | • | • | | | |
| Pharmacokinetics | | | | | | | • | | |
| Pharmacodynamics | | | | | | | | • | |
| QoL | | | | | | | | | • |

### 5.2.2 Subject Disposition

Subject disposition will be summarized for the Per-Protocol Analysis Set 1, Per-Protocol Analysis Set 2, and Full Analysis Set. All subjects who were screened for the study will be reported. The number of subjects who failed screening and the reasons for screen failures will be summarized.

The number of subjects who were randomized, treated, discontinued study treatment, and the reasons for discontinuation of study treatment will be summarized

Subject status (Alive, Death, Withdrew Consent, Lost to Follow-up, etc.) at the data cutoff for the primary analysis will also be summarized by treatment arm and overall.

### 5.2.3 Protocol Deviations

Protocol deviations will be identified and documented prior to database lock.

Protocol deviations will be classified as important or minor. Subjects who have experienced important protocol deviations will be excluded from per protocol analysis sets and reported in the clinical study report (CSR).

Important protocol deviations will be appropriately grouped into different categories and summarized by treatment arm.

The summary of important protocol deviations will be performed on Full Analysis Set. A list of subjects with important protocol deviations will also be provided.

### 5.2.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics will be summarized for the Per-Protocol Analysis Set 1, Per-Protocol Analysis Set 2, and Full Analysis Set.

The following demographic and baseline characteristics will be summarized:

- Age (years)
- Age group (<65 years, ≥65 years)
- Gender
- Race
- Race group (White, All Others)
- Ethnicity
- Region (Eastern Europe, Western Europe, North America, Asia/Pacific)
- Height (cm)
- Baseline weight (kg)
- Baseline body mass index (BMI) ($kg/m^2$)
- Karnofsky performance status (KPS)
- New York Heart Association (NYHA) classification
- Prior PD-1/PD-L1 treatment (Yes, No)
- MSKCC risk group (Favorable, Intermediate, and Poor risk)
- Baseline hypertension status (Yes, No)

The following disease history and characteristics at study entry will also be summarized:

- Time since the first RCC diagnosis to date of randomization (months)
- RCC diagnosis classification (predominant clear cell, other)

- Age at diagnosis (years)
- RCC American Joint Committee on Cancer (AJCC) Tumor-node-metastasis (TNM) staging at diagnosis
- Type of tumor lesions at screening (baseline):
  - Target lesions
    - Lymph node (Yes, No)
    - Non-lymph node (Yes, No)
  - Non-target lesions (Yes, No)

The following prior anticancer therapies will also be summarized:

- Number of prior therapy regimens
- Prior VEGF-targeted therapy (Yes, No)
- Prior biological agents (Yes, No)
- Duration of last therapy (months)
- Best response for last anticancer therapy (complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), not evaluable (NE), not applicable (NA), unknown)
- Time from end of last therapy to date of randomization (months)
- Prior treatment regimen (adjuvant, locally advanced, metastatic, neoadjuvant, unknown)
- Prior radiotherapy (Yes, No)
  - Time from last radiotherapy to date of randomization (months)
  - Site of prior radiotherapy
  - Tumor lesion at the site progressed since last radiotherapy (Yes, No, Not evaluated)
- Prior VEGF-targeted therapy
  - Duration of most recent VEGF-targeted therapy (months)
  - Best response for most recent VEGF-targeted therapy
  - Time from end of most recent VEGF-targeted therapy to date of randomization (months)
  - Type of prior VEGF-targeted therapy (adjuvant, locally advanced, metastatic, neoadjuvant, unknown)

In addition, a summary of stratification factors used in the randomization (MSKCC prognostic group and prior PD-1/PD-L1 treatment) based on IxRS data will be provided to evaluate whether or not randomization process was appropriately executed in the study.

Demographic and baseline disease characteristics, prior anticancer therapies will also be listed for each subject.

#### MEDICAL HISTORY

General medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 22.1 or higher and summarized by body system organ class (SOC) and preferred term (PT) for each treatment group and overall.

A subject data listing of medical history will be provided, including SOC and/or PT, current medical condition, date of diagnosis or surgical procedure or onset of symptoms, and end date/on-going.

### 5.2.5 Prior and Concomitant Therapy

All investigator terms for medications recorded on the case report form (CRF) will be coded using the World Health Organization Drug Dictionary (WHO DD) Version of March 2018 or later.

Prior medications will be defined as the medications that were started prior to the first dose of study drug. Concomitant medications will be defined as the medications that (i) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (ii) started on or after the date of the first dose of study drug up to 28 days following the last dose in randomization phase. Medications that cannot be determined to be prior/concomitant/posttreatment due to missing or incomplete dates will be regarded as concomitant.

Prior medications will be summarized by anatomical class (Anatomical Therapeutic Chemical [ATC] Level 1), pharmacologic class (ATC Level 3), and WHO DD preferred term. A similar summary will also be provided for concomitant medications except thyroxine suppression therapy and antihypertensive therapy. Concomitant thyroxine suppression therapy and antihypertensive therapy will be summarized separately. In addition, a separate summary for the concomitant medications of P-glycoprotein inhibitors and/or inducers will be provided.

Data listings will be generated for all prior and concomitant medications, concomitant thyroxine suppression therapy, concomitant antihypertensive therapy, and concomitant P-glycoprotein inhibitors and/or inducers.

Prior and concomitant medications/therapies will be summarized for Per-Protocol Analysis Set 1 and Full Analysis Set.

### 5.2.6 New Anticancer Therapies

The total number of subjects who received new anticancer therapies during Follow-up period will be reported for Per-Protocol Analysis Set 1 and Full Analysis Set in each treatment group. A summary of new anticancer therapy will be presented by anatomical class (ATC Level 1), pharmacologic class (ATC Level 3), and WHO DD preferred term. The number of anticancer regimens received, the duration of the first new anticancer regimen, and the primary reason for discontinuation of the first new anticancer regimen will also be summarized.

The new anticancer procedures during Follow-up period will also be reported.

### 5.2.7 Treatment Compliance

Records of treatment compliance for each subject will be kept during the study, including the Follow-up Period. Clinical research associates will review treatment compliance during investigational site visits and at the completion of the study. Received dose as percent of planned dose per subject (ie, relative dose intensity) will be summarized as described in Section 5.6.2.2, no other analysis of treatment compliance is planned.

## 5.3 Data Analysis General Considerations

### 5.3.1 Pooling of Centers

Subjects from all centers will be pooled for analyses. For the subgroup analysis for region, subjects from all centers within each region will be pooled for the analysis.

### 5.3.2 Adjustments for Covariates

For the analyses of primary efficacy endpoint of $ORR_{24W}$ and primary safety endpoint of proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization, the CMH test stratified by the randomization stratification factors of MSKCC prognostic group and prior PD-1/PD-L1 treatment will be used.

The stratified analyses will be performed using the stratification factors from IxRS data. For the stratification factor of MSKCC prognostic group from IxRS, some subjects with MSKCC Prognostic Score 2 were misclassified to intermediate risk on IRT system which should have been classified to poor risk per protocol. The poor risk group will be pooled with the intermediate risk group in the relevant stratified analyses based on IxRS data, the favorable risk group will remain intact. The analyses based on the stratification factors from CRF data in clinical database may also be performed as sensitivity analyses.

To explore homogeneity of treatment effect across centers/regions, sensitivity analyses will be conducted to adjust center/region effect for the primary endpoints of $ORR_{24W}$ and the proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after

randomization using logistic regression models including center/region and stratification factors as covariates.

Other demographic and baseline variables may be included and adjusted in the statistical models as sensitivity analyses if deemed necessary.

### 5.3.3 Multiple Comparisons/Multiplicity

There will be two interim analyses and one final analysis for the non-inferiority test on the primary efficacy endpoint $ORR_{24W}$. O'Brien-Fleming efficacy boundary is used to control the overall Type I error rate at the nominal level ($\alpha$=0.05 one-sided). Specifically, if the 1-sided *P* value is ≤0.005 at the first interim analysis, or ≤0.014 at the second interim analysis, or ≤0.045 at the final analysis, non-inferiority in $ORR_{24W}$ will be claimed.

A non-binding futility will be also carried out at each interim analysis. If the 1-sided *P* value is ≥0.776 at the first interim analysis, or ≥0.207 at the second interim analysis, futility will be claimed.

These stopping boundaries are based on the plan that the two interim analyses will take place when the total number of 150 subjects and 200 subjects in the Per-protocol Analysis Set 1 have completed 24 weeks of follow-up or discontinued early.

Superiority on the primary safety endpoint will be tested at 2-sided $\alpha$=0.05 only if the non-inferiority on the primary efficacy endpoint $ORR_{24W}$ is claimed.

All secondary efficacy endpoints will be analyzed descriptively, no formal statistical hypothesis tests will be performed on the secondary efficacy endpoints.

### 5.3.4 Examination of Subgroups

In general, subgroup analyses on the prespecified subgroups in Table 4 will be performed for the efficacy endpoint ORR based on the investigator assessment and the primary safety endpoint of proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization.

Exploratory analysis on the subgroup of subjects previously treated with PD-1/PD-1 will be performed and is specified in the Section 5.8.2.

Additional subgroup analyses may be performed as appropriate.

**Table 4: Subgroup Analyses for Efficacy and Safety Endpoints**

| Subgroup | Definition |
|---|---|
| Sex | Male, Female |
| Age | <65, ≥65 years |

| | |
|---|---|
| Race | White, Non-White |
| Region | Eastern Europe, Western Europe, North America, Asia/Pacific |
| Baseline hypertension status | Yes, No |
| MSKCC risk group | Favorable, Intermediate, and Poor risk |
| Prior PD-1/PD-L1 treatment | Yes, No |

### 5.3.5 Handling of Missing Data, Dropouts, and Outliers

For efficacy endpoints related to ORR which summarize the percentage of responders, missing responses will not be imputed and subjects with missing responses will be considered as non-responders.

For incomplete dates involving efficacy and safety data such as adverse events, concomitant medications, laboratory assessments, vital signs, and echocardiogram/multigated acquisition (MUGA) data, a conservative imputation will be used for calculation if needed. The imputation rules will be specified in study analysis dataset specification with more details.

HRQoL analysis will follow FDA and EMEA PRO guidelines. Handling of missing values for HRQoL analysis will be specified in details in a separate SAP and HRQoL report.

### 5.3.6 Other Considerations

There are no other specific considerations that are not covered in other sections of this SAP.

## 5.4 Efficacy Analyses

Tumor response will be assessed using RECIST 1.1 criteria (Eisenhauer, 2009). For this study, a confirmation of CR or PR is required by a consecutive tumor assessment no less than 4 weeks after the initial assessment of response.

The primary efficacy analyses will be based on the investigator assessments. The stratification factors (ie, MSKCC prognostic group and prior PD-1/PD-L1 treatment) from IxRS data will be used for the primary analysis. For the stratification factor of MSKCC prognostic group from IxRS, some subjects with MSKCC Prognostic Score 2 were misclassified to intermediate risk on IRT system which should have been classified to poor risk per protocol. The poor risk group will be pooled with the intermediate risk group in the relevant stratified analyses based on IxRS data, the favorable risk group will remain intact.

Sensitivity analyses may be performed using the stratification factors from CRF data in clinical database.

Exploratory analyses for the primary efficacy endpoint $ORR_{24w}$ and secondary efficacy endpoints of ORR and PFS based on IIR assessments will also be performed (refer to Section

5.8.1).

### 5.4.1 Primary Efficacy Analyses

The primary efficacy endpoint $ORR_{24W}$ is defined as the proportion of subjects with BOR of CR or PR as of the Week 24 (after randomization) timepoint based on investigator assessments, during treatment or within 28 days after the last dose date but on or prior to the start of new anticancer therapy. Subjects with responses achieved after the start of new anticancer therapy or after 28 days from the last dose date will not be considered as responders.

The primary analysis of $ORR_{24W}$ will be based on a non-inferiority test. The study hypothesis is that the $ORR_{24W}$ for lenvatinib 14 mg QD in combination with everolimus 5 mg is non-inferior to the $ORR_{24W}$ for lenvatinib 18 mg QD in combination with everolimus 5 mg in subjects with renal cell carcinoma following one prior VEGF-targeted treatment.

For this study, the non-inferiority of lenvatinib 14 mg QD in combination with everolimus 5 mg relative to lenvatinib 18 mg QD in combination with everolimus 5 mg is defined using 70% retention of the treatment effect of lenvatinib 18 mg QD in combination with everolimus vs. everolimus, which translated to the estimated odds ratio =0.76 for the non-inferiority margin following the approach of Rothmann et al. (2003).

The non-inferiority hypothesis for $ORR_{24W}$ with the non-inferiority margin of odds ratio =0.76 for the final analysis can be stated as follows:

$H_0$: Odds ratio of $ORR_{24W\text{-}14mg}$ vs $ORR_{24W\text{-}18mg} \leq 0.76$

$H_a$: Odds ratio of $ORR_{24W\text{-}14mg}$ vs $ORR_{24W\text{-}18mg} > 0.76$ (non-inferiority)

The primary analysis of $ORR_{24W}$ will be performed on the Per-Protocol Analysis Set 1. The point estimate of $ORR_{24W}$ with the corresponding 95% CI for each treatment arm will be summarized. The odds ratio of $ORR_{24W}$ (lenvatinib 14 mg over lenvatinib 18 mg) along with the 90% CI will be calculated using the CMH method stratified by MSKCC prognostic group and prior PD-1/PD-L1 treatment. Non-inferiority in $ORR_{24W}$ will be claimed if the O'Brien-Fleming efficacy boundary is crossed. If the 1-sided *P* value is ≤0.045 at the final analysis, non-inferiority in $ORR_{24W}$ will be claimed (interim analysis is specified in Section 6). The boundary *P* value ≤0.045 for the final analysis is based on the plan that the two interim analyses will take place when the total number of 150 subjects and 200 subjects in the Per-Protocol Analysis Set 1 have completed 24 weeks of follow-up or discontinued early.

The treatment difference in $ORR_{24W}$ between lenvatinib 14 mg and lenvatinib 18 mg will also be estimated along with 90% CIs based on the asymptotic normal approximation.

If non-inferiority in $ORR_{24W}$ is claimed, the superiority test at 2-sided $\alpha$=0.05 for the primary safety endpoint will be performed. The details are specified in Primary Safety Endpoint Section

5.6.1.

Sensitivity analyses of $ORR_{24W}$ based on Full Analysis Set and Per-Protocol Analysis Set 2 will be performed in a similar manner as described above.

To explore homogeneity of treatment effect across centers/regions, a sensitivity analysis of $ORR_{24W}$ using logistic regression model with center/region along with stratification factors as covariates will be conducted to adjust center/region effect.

Sensitivity analyses of $ORR_{24W}$ based on the stratification factors from CRF data in clinical database will also be performed on Per-Protocol Analysis Set 1, Full Analysis Set and Per-Protocol Analysis Set 2.

For this study, for subjects in the lenvatinib 14 mg QD with everolimus 5 mg arm, lenvatinib dose should be escalated to 18 mg QD (with everolimus 5 mg) as long as there are no intolerable Grade 2 or any ≥ Grade 3 TEAEs that require dose reductions in the first 28-day cycle. To evaluate any potential impact due to failing to up-titrate the doses for eligible subjects in the lenvatinib 14 mg QD with everolimus 5 mg arm, another sensitivity analysis may be performed on Per-Protocol Analysis Set 1 to exclude those subjects who should have been up-titrated by protocol but were actually not up-titrated by investigators, if the number of such cases in the lenvatinib 14 mg QD with everolimus 5 mg arm is deemed high (eg, ≥10%).

An analysis of $ORR_{24W}$ based on the data prior to when the protocol was amended to open-label (from double-blinded) (ie, treatment assignments were unblinded to investigator sites on 14 July 2018) will be performed on Per-Protocol Analysis Set 1 and Per-Protocol Analysis Set 2 to check for consistency of study results before and after the protocol became un-blinded.

### 5.4.2 Secondary Efficacy Analyses

The secondary efficacy endpoints include PFS, ORR, OS, PFS2. The analysis of secondary efficacy endpoints will be performed descriptively based on the Per-Protocol Analysis Set 1 as the primary analysis set, Per-Protocol Analysis Set 2 and Full Analysis Set as the secondary analysis sets.

#### 5.4.2.1 Progression-free Survival

DEFINITION

PFS is defined as the time from the date of randomization to the date of first documentation of disease progression based on investigator assessments or date of death, whichever occurs first.

Determination of the date of PFS event or censoring is summarized in Table 5 below.

**Table 5 PFS Event and Censoring Rules**

| No. | Situation | Date of Progression or Censoring | Outcome |
|---|---|---|---|
| 1 | No baseline or postbaseline adequate tumor assessments | Date of randomization | Censored |
| 2 | Progression documented between scheduled visits, on or prior to new anticancer therapy[#] and within 28 days after the last dose of study treatment | Date of first radiologic PD assessment | PFS Event |
| 3 | No progression at the time of data cutoff | Date of last adequate radiologic assessment on or prior to data cutoff | Censored |
| 4 | New anticancer therapy started | Date of last adequate radiologic assessment on or prior to the start of new anticancer therapy | Censored |
| 5 | Death before first PD assessment, on or prior to start of new anticancer therapy[#] and within 28 days after the last dose of study treatment | Date of death | PFS Event |
| 6 | Death between adequate assessment[#,*] | Date of death | PFS Event |
| 7 | Death or progression after more than one missed visit/tumor assessment** | Date of last adequate radiologic assessment before missed tumor assessments | Censored |
| 8 | Death or progression after 28 days from the last dose of study treatment | Date of last adequate radiologic assessment before the date of last dose of study treatment + 28 days | Censored |
| 9 | No progression and treatment discontinuation for reasons other than PD | Date of last adequate radiologic assessment before treatment discontinuation (ie, within last dose date + 28 days) | Censored |

Note:

# If documented PD and/or death occurs after the last dose, it is counted as a PFS event as long as the PD and/or the death occurs within 28 days since the date of the last dose of study treatment and provided that it does not violate other censoring rules (eg, start of new anticancer therapy before the PD or death). Otherwise the subject will be censored on the date of the last tumor assessment before the date of last dose of the study treatment.

* 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 therapy starts or after 28 days from the last dose date will not be considered in the definition of PFS.

** More than one missed visit/adequate tumor assessment is defined as having the duration between two consecutive tumor assessments or the duration between the last adequate tumor assessment and death/PD being longer than (>) 118 days (the visit schedule for tumor assessment is during every 8$^{th}$ week for this study).

The priority of the censoring rules is described as follows:

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

- If a subject did not have baseline or postbaseline adequate tumor assessments (No. 1), the subject will be censored at the date of randomization. However, if the subject died within 118 days after randomization and did not receive any new anticancer therapy, it will be counted as a PFS event at the date of death. If a subject had new anticancer therapy before PD or death (No. 4), the subject will be censored at the date of the last adequate tumor assessment on or prior to the date of new anticancer therapy. If PD is reported after the date of the last dose + 28 days, the subject will be censored at the date of last tumor assessment before the date of last dose + 28 days (No. 8). If PD or death happened within 28 days after the last dose date, it will be considered as a PFS event.
- If a subject missed two or more tumor assessments before PD or death (No. 7), the subject will be censored at the date of the last adequate tumor assessment before the missed tumor assessments. Note that if a subject is censored by both of this criterion and start of new anticancer therapy criterion, the earliest censoring date will be used.
- Otherwise, if a subject had PFS 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).

ANALYSIS METHOD

The analysis of PFS will be performed based on the Per-Protocol Analysis Set 1 as the primary analysis set. The median PFS and the PFS rates at 6, 12, 18 months, and so on (depending on data adequacy) will be estimated using the Kaplan–Meier method and presented with 2-sided 95% CIs. The hazard ratio of lenvatinib 14 mg over lenvatinib 18 mg and the corresponding 90% CIs will be estimated using the Cox regression model with Efron's method for ties, stratified by MSKCC prognostic group and prior PD-1/PD-L1 treatment. The Kaplan-Meier curve of PFS will also be plotted over time for each treatment arm.

Sensitivity analyses of PFS based on the Per-Protocol Analysis Set 2 and Full Analysis Set will be performed in a similar manner as described above.

#### 5.4.2.2 Objective Response Rate

ORR is defined as the proportion of subjects achieving a best overall response of confirmed CR or PR during treatment or within 28 days after the last dose date but on or prior to the start of new anticancer therapy at the time of clinical data cutoff. Subjects with overall responses achieved after the start of new anticancer therapy or after 28 days from the last dose date will not be considered as responders. Subjects who do not have a tumor response assessment for any reason will be considered non-responders and will be included in the denominator when calculating the response rate.

The point estimate of ORR and corresponding 95% CI based on asymptotic normal approximation will be summarized for each treatment arm. Odds ratio with 90% CI will be estimated using the CMH method stratified by MSKCC prognostic group and prior PD-1/PD-L1 treatment. In addition, the treatment difference in ORR between lenvatinib 14 mg and lenvatinib 18 mg will also be estimated along with 90% CIs based on the asymptotic

normal approximation.

A sensitivity analysis of ORR based on the Per-Protocol Analysis Set 2 and Full Analysis Set will be performed in a similar manner as described above.

In addition, waterfall plot for the percentage changes from baseline to postbaseline nadir (ie, best percentage changes from baseline) in sum of diameters of the target lesions will be presented. The depth of response for >0%, ≥30%, ≥50% and ≥75% maximum reduction in sum of diameters will be summarized.

#### 5.4.2.3 Overall Survival

Overall survival is defined as the time from the date of randomization to the date of death. All deaths during the study will be considered as OS events.

Subjects who were lost to follow-up or withdrew consent will be censored at the date the subject was last known to be alive. Subjects who are still alive at the clinical cut-off date will be censored at the date of data cutoff.

Determination of date of OS event or censoring is summarized in Table 6 below.

**Table 6 OS Event and Censoring Method**

| Situation | Date of Event or Censoring | Outcome |
|---|---|---|
| Death during study | Date of death | OS Event |
| Death after data cut-off | Date of data cut-off | Censored |
| Subject still alive at data cut-off | Date of data cut-off | Censored |
| Subject lost to follow-up or withdrawal of consent before data cut-off | Date of last known to be alive | Censored |

The analysis of OS will be performed on the Per-Protocol Analysis Set 1 as the primary analysis set. The median OS and the OS rates at 12, 18, and 24 months will be estimated using the Kaplan–Meier method and presented with 2-sided 95% CIs. The hazard ratio of lenvatinib 14 mg over lenvatinib 18 mg and the corresponding 90% CIs will be estimated using the Cox regression model with Efron's method for ties, stratified by MSKCC prognostic group and prior PD-1/PD-L1 treatment. The Kaplan-Meier curve of OS will also be plotted over time for each treatment arm.

Sensitivity analyses of OS based on the Per-Protocol Analysis Set 2 and Full Analysis Set will be performed in a similar manner as described above.

#### 5.4.2.4 Progression-free Survival on Next Line of Therapy

PFS2 is defined as the time from randomization to the date of disease progression on next line of therapy or death, whichever comes first.

Subjects who started the next line of therapy and have not yet progressed on the next line of therapy and are still alive will be censored on the last date they are known to be alive. Subjects who did not have any new anticancer therapies and had no death documented before the data cutoff date will be censored at the last date they are known to be alive.

Determination of date of PFS2 event or censoring is summarized in Table 7 below.

**Table 7 PFS2 Event and Censoring Method**

| Situation | Date of Event or Censoring | Outcome |
|---|---|---|
| Subjects who started the next line of therapy and have not yet progressed on the next line of therapy and are still alive | Date of last known to be alive | Censored |
| Subjects who did not have any new anticancer therapies and had no death documented before the data cutoff date | Date of last known to be alive | Censored |
| Disease progression on the next line of therapy or any death | Minimum of earliest date that indicates progression on the next line of therapy and date of death | PFS2 event |
| Other, such as:<br>o Lost to follow-up<br>o Withdrawal of consent | Date of last known to be alive | Censored |

The analysis of PFS2 based on investigator assessment will be performed on Per-Protocol Analysis Set 1. The median PFS2 will be estimated using the Kaplan–Meier method and presented with 2-sided 95% CIs. The hazard ratio of lenvatinib 14 mg over lenvatinib 18 mg and the corresponding 90% CIs will be estimated using the Cox regression model with Efron's method for ties, stratified by MSKCC prognostic group and prior PD-1/PD-L1 treatment. The Kaplan-Meier curve of PFS2 will also be plotted for each treatment arm.

Sensitivity analyses of PFS2 based on the Per-Protocol Analysis Set 2 and Full Analysis Set will also be performed in a similar manner.

### 5.4.3 Other Efficacy Analyses

Other efficacy analyses may be performed descriptively on duration of response (DOR), disease control rate (DCR), and clinical benefit rate (CBR).

#### 5.4.3.1 Duration of Response

Duration of response is calculated from the date of initial documentation of the best overall response of CR or PR to the date of first objectively documented progressive disease or death, whichever occurs first, for the subjects who had achieved a best overall response of CR or PR.

Subjects who start new anticancer therapies without PD will be censored at the date of the last available tumor assessment prior to the start of new anticancer therapies. Subjects who have not progressed and subjects who did not die will be censored at the last tumor assessment date.

Analysis of DOR will be performed on the subjects who achieved a best overall response of CR or PR. Median DOR with 2-sided 95% CI will be estimated using the Kaplan-Meier method.

#### 5.4.3.2 Disease Control Rate

The DCR is the proportion of subjects who achieved a best overall response of CR, PR or SD (minimum duration of SD ≥7 weeks). The DCR with corresponding 2-sided 95% CI based on the asymptotic normal approximation will be provided.

#### 5.4.3.3 Clinical Benefit Rate

The CBR is the proportion of subjects who achieved a best overall response of CR, PR or durable SD (duration of SD ≥23 weeks). The CBR with corresponding 2-sided 95% CI based on the asymptotic normal approximation will be provided.

### 5.4.4 Subgroup Analysis for Efficacy/Safety Endpoints

To assess the internal consistency and investigation of homogeneity of the treatment effect across subgroups, subgroup analyses for the efficacy endpoint of ORR based on the investigator assessment and primary safety endpoint of proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization based on predefined subgroups specified in Section 5.3.4 will be conducted.

Forest plots of subgroup analyses will be generated.

Additional subgroup analyses may be performed for selected efficacy and/or safety endpoints.

## 5.5 Pharmacokinetic, Pharmacodynamic, and Other Biomarker Analyses

Analyses related to PK, PD, and biomarkers will be detailed in separate analysis plans.

## 5.6 Safety Analyses

The primary safety endpoint of the study is the proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization. The primary safety endpoint will be performed on the Per-Protocol Safety Analysis Set, Safety Analysis Set will be used for descriptive purpose. A sensitivity analysis for primary safety endpoint based on Per-Protocol Analysis Set 2 will also be performed. Other safety assessments include TEAEs, clinical laboratory evaluations, vital signs, electrocardiograms (ECGs), and echocardiogram or MUGA scan including left ventricular ejection fraction (LVEF). All safety analyses will be performed on the Safety Analysis Set. Some selected safety analyses will also be performed on Per-Protocol Safety Analysis Set.

### 5.6.1 Primary Safety Endpoint

If the boundary for the non-inferiority test on the primary efficacy endpoint $ORR_{24W}$ is crossed, the following superiority test at 2-sided $\alpha$=0.05 on the primary safety endpoint (ie, the proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization) will be performed:

$H_0$: $\Delta_{14mg\text{-}18mg} = 0$

$H_a$: $\Delta_{14mg\text{-}18mg} \neq 0$

where $\Delta_{14mg\text{-}18mg}$ is the difference in the proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization between lenvatinib 14 mg and lenvatinib 18 mg arm.

The proportion of subjects with intolerable Grade 2 or any ≥ Grade 3 TEAEs within 24 weeks after randomization will be tested using the CMH test at 2-sided $\alpha$=0.05, stratified by MSKCC prognostic group and prior PD-1/PD-L1 treatment from IxRS data.

At each interim analysis and the final analysis, if the non-inferiority boundary of the primary efficacy endpoint is crossed, the analysis of the primary safety endpoint will be performed on Per-Protocol Safety Analysis Set. The proportions will also be summarized on Safety Analysis Set for descriptive purpose. A sensitivity analysis on primary safety endpoint will be performed on Per-protocol Analysis Set 2.

A sensitivity analysis for the primary safety endpoint based on the stratification factors from CRF data in clinical database will also be performed on Per-Protocol Safety Analysis Set.

To explore homogeneity of treatment effect across centers/regions, a sensitivity analysis for the primary safety endpoint using logistic regression model with center/region along with stratification factors as covariates will be conducted to adjust center/region effect.

The summary of intolerable Grade 2 and ≥ Grade 3 TEAEs by SOC and PT is specified in Section 5.6.3.3.

### 5.6.2 Extent of Exposure

Extent of exposure to study treatment, study drug administered, and study drug modifications will be summarized and presented for each individual drug and overall. The dosing and treatment exposure for those 32 subjects who had received ≥2 incorrect lenvatinib doses due to IxRS issue will also be summarized and listed.

#### 5.6.2.1 Extent of Exposure to Study Treatment

Treatment duration and the total number of treatment cycles will be summarized descriptively. The maximum number of treatment cycles for each subject is the largest cycle number in which a subject receives any non-zero dose of study drugs. The number and percentage of subjects treated within each cycle will also be summarized for each treatment arm.

Treatment duration will be summarized descriptively for each individual drug and overall. 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 each individual drug in months will be calculated as (Date of the last non-zero dose – Date of the first non-zero dose + 1)/30.4375 for the specific drug, 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.

In this study, study drugs refer to lenvatinib 14 mg, lenvatinib 18 mg, and everolimus 5 mg.

#### 5.6.2.2 Study Drug Administered

The total dose received per subject, dose intensity, and relative dose intensity will be summarized with descriptive statistics for each individual study drug.

For each individual study drug, the total dose per subject (mg) will be calculated as the sum of all doses that subject has received during the study. The dose intensity (mg/day) will be calculated as the total dose (mg) received during the study divided by the treatment duration of the specific drug (days). The relative dose intensity, which is defined as the ratio (%) of total dose received and total planned dose (ie, received dose as a percentage of planned dose), will be calculated as the dose intensity divided by the daily planned dose.

#### 5.6.2.3 Study Drug Dose Reductions and Interruptions

DOSE REDUCTIONS

Dose reduction refers to a situation that a dose level was reduced from the previous dose level without going back. Dose reduction could apply to any of the study drugs: lenvatinib18 mg/day (could reduce to 14, 10, 8, and 4 mg/day), lenvatinib 14 mg/day (could reduce to 10, 8, and 4 mg/day), and everolimus (5 mg every day could reduce to 5 mg every other day).

DOSE INTERRUPTIONS

Dose interruption refers to a situation that a subject had a planned temporary break from taking the study drug for a short period (ie, interruption period). Dose interruption could apply to any of the study drugs.

Dose interruption only refers to the scenario that the dose levels or dosing frequencies before and after interruption period (defined as the period with dose=0) are the same. For example: 18 mg lenvatinib followed by 0 mg and followed by 18 mg lenvatinib; 5 mg everolimus every day followed by 0 mg followed by 5 mg everolimus every day. If the dose level after dose interruption period was reduced from the dose level before the interruption period, it should be counted as dose reduction and should not be counted as dose interruption. The following 2 scenarios are examples of dose reduction instead of dose interruption:

(1) 18 mg lenvatinib followed by 0 mg followed by 14 mg lenvatinib;

(2) 5 mg everolimus every day followed by 0 mg followed by 5 mg everolimus every other day;

Please note, for the 2 scenarios above, the period with dose=0 should be counted as dose reduction period and not dose interruption (ie, the dose reduction date should be the date of dose=0). If the subject discontinued from treatment permanently after dose interruption with dose=0, it should be counted as treatment discontinuation instead of dose interruption.

The number of subjects with dose reductions and dose interruptions for each individual study drug will be summarized. Number (ie, frequency) of dose reductions per subject will also be summarized.

Time to first dose reduction is defined as the time period from the first dose date to the date of first dose reduction, or the time period from the first dose date to the date of first dose interruption for those subjects who had dose interruption first and then followed by dose reduction. Time to first dose reduction will be derived and summarized descriptively for each individual drug and overall for those subjects who had dose reduction during treatment period.

Median time to first dose reduction with 95% CI will also be estimated based on Kaplan-Meier method for all the treated subjects. Subjects who did not have dose reductions during treatment period will be censored at the date of last dose.

### 5.6.3 Adverse Events

Adverse events will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE), Version 4.03. The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the MedDRA Version 22.1 or higher. Adverse events will be coded to the MedDRA lower level term (LLT) closest to the verbatim term, the linked MedDRA preferred term and primary system organ class are also captured in the database.

A treatment-emergent AE is defined as any one of the followings:

- An AE that emerges during treatment or up to 28 days after the last dose of study treatment, having been absent at pretreatment (baseline); or
- An AE that reemerges during treatment or up to 28 days after the last dose of study treatment, having been present at pretreatment (baseline) but stopped before treatment; or
- An AE that worsens in toxicity grade during treatment or up to 28 days after the last dose of study treatment, relative to the pretreatment state, when the AE is continuous; or
- An AE with missing or partial onset date and is considered related to study treatment by investigator; or
- An AE that is not treatment-related and with missing or partial onset date will be considered as treatment-emergent unless the onset of the AE can be determined as earlier than the first dose date or later than 28 days after the last dose of study treatment.

Unless otherwise specified, a subject will be counted only once at each level of summary (eg, SOC and/or PT), even if the subject experienced more than 1 TEAE within a specific SOC and/or PT.

#### 5.6.3.1 Overview of TEAEs

An overview of TEAEs reported through the study will be provided for each treatment arm. Overall summary of TEAE will include the number and percentage of subjects with TEAEs, serious TEAEs, treatment-related TEAEs, intolerable Grade 2 TEAEs, TEAEs of Grade 3 or higher, TEAEs leading to treatment discontinuation (discontinuation of both study drugs in each treatment arm), TEAEs leading to dose reductions, TEAEs leading to dose interruptions, and TEAEs with fatal outcomes. Overview of TEAEs episodes adjusted by treatment duration will also be provided.

An overview of TEAEs reported for those 32 subjects who had received ≥2 incorrect lenvatinib doses due to IxRS issue will also be provided.

#### 5.6.3.2 All TEAEs

The following summaries will be provided for all TEAEs:

- TEAEs by SOC and PT
- TEAEs by decreasing frequency of PT
- Most common (eg, ≥10%) TEAEs by SOC and PT
- TEAEs by SOC, PT, and worst toxicity grade

#### 5.6.3.3 Intolerable Grade 2 and ≥ Grade 3 TEAEs

The intolerable Grade 2 and ≥ Grade 3 TEAEs will be summarized by SOC and PT.

In addition, a listing of intolerable Grade 2 and ≥ Grade 3 AEs will be provided.

#### 5.6.3.4 Treatment-related TEAEs

The following treatment-related TEAEs will be summarized:

- Treatment-related TEAEs by SOC, PT and worst toxicity grade
- Most common (eg, ≥10%) treatment-related TEAEs by SOC and PT

#### 5.6.3.5 Serious TEAEs

The incidence of serious TEAEs will be summarized as below:

- Serious TEAEs by SOC and PT
- Treatment-related serious TEAEs by SOC and PT
- Serious TEAEs by decreasing frequency of PT
- Most commonly reported (eg, ≥5%) serious TEAEs by SOC and PT
- Serious TEAEs adjusted by treatment duration

In addition, a listing of serious AEs will be provided.

#### 5.6.3.6 TEAEs Leading to Treatment Discontinuation

The proportion of subjects who discontinued treatment due to toxicity and time to treatment discontinuation due to toxicity will be summarized descriptively.

The TEAEs leading to treatment discontinuation (discontinuation of both study drugs in each treatment arm) will be summarized by SOC and PT for those subjects indicated as having discontinued study treatment due to an adverse event on the CRF treatment disposition page.

The TEAEs leading to study drug discontinuation for each individual drug will also be summarized.

A listing of AEs leading to treatment discontinuation or any study drug discontinuation will be provided.

#### 5.6.3.7 TEAEs Leading to Dose Reductions and Dose Interruptions

The TEAEs leading to dose reductions and dose interruptions will be summarized by SOC and PT for each treatment arm. The TEAEs leading to dose reductions and dose interruptions will be summarized for each individual drug and overall.

A listing of AEs leading to dose reductions and/or dose interruptions of any study drug will be provided.

#### 5.6.3.8 Grade 3 or Higher TEAEs

The following Grade 3 or higher TEAEs will be summarized:

- Grade 3 or higher TEAEs by SOC and PT
- Most commonly reported (eg, ≥5%) Grade 3 or higher TEAEs by SOC and PT
- Grade 3 or higher TEAEs leading to treatment discontinuations, dose reductions and/or dose interruptions by SOC and PT

In addition, a listing of Grade 3 or higher AEs will be provided.

#### 5.6.3.9 TEAEs with Fatal Outcome

The TEAEs with fatal outcome will be summarized by SOC and PT for each treatment arm. Treatment-related TEAEs with fatal outcome will also be summarized.

A listing of AEs with fatal outcome will be provided.

#### 5.6.3.10 Deaths

The number of subjects who died during the study will be summarized.

A listing of subjects who died during the study will be provided.

#### 5.6.3.11 Adverse Events of Clinical Interest

Clinically significant adverse events (CSAEs) for lenvatinib and treatment-emergent AEs of special interest (AEoSI) for everolimus will be identified based on a thorough review of safety data from the clinical and pharmacovigilance database.

Evaluations of the CSAEs and AEoSI will be based on standardized MedDRA queries (SMQs), customized MedDRA queries (CMQ), or sponsor-generated queries (SGQs), which included sponsor-specified PT either alone or in addition to those listed for an SMQ or CMQ.

CSAEs and AEoSI will be summarized by overall, CTCAE grade and treatment cycle. The subjects with CSAEs and AEoSI leading to dose reductions, dose interruptions, and treatment discontinuations will also be provided.

In addition, number of subject with hypertension leading to administration of concomitant medication and type of antihypertensive medication received will be summarized. Time to first onset of hypertension (defined as the time from the date of first dose date to the date of first onset of hypertension) will also be summarized. Similar analysis will also be provided for the subjects who had proteinuria.

Hypertension based on vital sign data will also be summarized. Details are specified in the vital sign Section 5.6.5.

Proteinuria based on urine dipstick data will also be summarized. Details are specified in the urinalysis Section 5.6.4.2.

### 5.6.4 Laboratory Values

Laboratory results will be summarized using Système International (SI) units. 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. Central laboratory test results will be used as the 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.

On-treatment laboratory tests will be defined as the laboratory tests conducted from the start of treatment to no more than 28 days after the last dose of study treatment.

#### 5.6.4.1 Hematology and Clinical Chemistry

Laboratory parameters will be graded based on CTCAE Version 4.03. In the summary of laboratory parameters by CTCAE grade, for parameters with CTCAE grading in both high and low direction (eg, calcium, glucose, magnesium, potassium, sodium), CTCAE grades in both high and low directions will be summarized separately.

The evaluation of clinical laboratory tests will focus on the following selected laboratory analytes:

Hematology Panel:

- hematocrit, hemoglobin, red blood cell (RBC) counts, white blood cell (WBC) count, absolute neutrophil count, and absolute lymphocyte count, platelet count

Chemistry Panel:

- Electrolytes: bicarbonate, calcium, chloride, magnesium, phosphate, potassium, sodium
- Liver function tests: alanine aminotransferase (ALT), alkaline phosphatase, aspartate aminotransferase (AST), direct bilirubin, total bilirubin
- Renal function tests: blood urea/blood urea nitrogen (BUN), creatinine
- Thyroid function tests: thyroid stimulating hormone (TSH), free T4 level
- Other: albumin, cholesterol, triglycerides, glucose, calcium, lactate dehydrogenase (LDH), amylase, lipase, creatine kinase

Descriptive statistics for values and changes from baseline at each scheduled visit for hematology and chemistry laboratory parameters will be provided. Box plots of laboratory values over time will also be provided for selected hematology and chemistry parameters.

In addition, the worst CTCAE grade in hematology and chemistry during the treatment will be summarized by treatment arm and CTCAE grade. Shift tables from baseline to the worst CTCAE grade during treatment will be generated.

Subjects with treatment-emergent markedly abnormal values (TEMAV) will be summarized. Treatment-emergent markedly abnormal value is defined as a value that is outside of 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.

#### 5.6.4.2 Urinalysis

Proteinuria: Shifts from baseline to worst postbaseline for proteinuria determined by dipstick (negative, trace, 1+, 2+, 3+, 4+) will be presented by treatment arm.

The urinalysis results from 24-hour urine protein will also be summarized by worst Grade.

### 5.6.5 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure, pulse, respiratory rate, temperature, and weight) and changes from baseline at each scheduled visit will be presented.

Blood pressure will also be summarized using a shift table from baseline to worst postbaseline CTCAE grade by the categories defined in Table 8 below based on CTCAE grades.

**Table 8 Blood Pressure 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 |

Subjects with systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg during treatment will be listed along with their hypertensive medications.

### 5.6.6 Electrocardiograms

Descriptive statistics for ECG parameters and changes from baseline at each scheduled visit will be presented.

Shift tables will be presented for the changes from baseline in ECG findings (categorized as normal; abnormal, not clinically significant; and abnormal clinically significant) for each scheduled visit.

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

- Absolute QTc interval prolongation:
  - QTc interval >450 msec
  - QTc interval >480 msec
  - QTc interval >500 msec
- Change from baseline in QTc interval:
  - QTC interval increases from baseline >30 msec
  - QTC interval increases from baseline >60 msec

A subject listing will also be provided.

### 5.6.7 Other Safety Analyses

#### 5.6.7.1 Karnofsky Performance Status

Karnofsky performance status (KPS) will be summarized by a shift table from baseline to worst postbaseline scale.

#### 5.6.7.2 Left Ventricular Ejection Fraction

LVEF (%) assessed on echocardiogram or MUGA scans will be summarized. The lowest postbaseline value and change from baseline will be summarized descriptively.

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 may be generated.

## 5.7 Other Analyses

### 5.7.1 Health-Related Quality of Life

The analysis of HRQoL will be performed on the QoL Analysis Set.

For HRQoL analysis, summary statistics of the scores for the derived functional/symptom scales according to the scoring manual and global health status scores will be summarized by treatment arm at each time point. A separate prespecified HRQoL analysis following FDA and EMEA PRO Guidelines will be performed and detailed in a separate SAP and HRQoL report. Scoring of EQ-5D-3L and derivation of utility for health economic analysis will also be accomplished in a separate analysis and described in a separate HRQoL report.

## 5.8 Exploratory Analyses

### 5.8.1 Efficacy Analysis Based on IIR Assessment

The efficacy endpoints $ORR_{24W}$, ORR, and PFS based on blinded IIR assessment will be defined and analyzed in the same manner as those based on the investigator assessments that described in Section 5.4.

### 5.8.2 Exploratory Analysis for Subjects with Prior PD-1/PD-L1 Treatments

Exploratory analysis on efficacy endpoints of PFS and OS may be performed descriptively for the subjects who had prior PD-1/PD-L1 treatments.

The subgroup analyses on the efficacy endpoint ORR and primary safety endpoint for the subjects with prior PD-1/PD-L1 treatment are specified in Section 5.4.4.

### 5.8.3 Biomarker-related Analysis

For the biomarker-related exploratory analysis, such as the analyses to explore the associations between blood biomarker and efficacy-related endpoints and the development of exposure/biomarker/clinical endpoint models (whenever possible, using a mechanism-based approach) for both efficacy and safety data, the analysis methods will be specified in a separate analysis plan.

Other exploratory analyses may be conducted as appropriate.

# 6 INTERIM ANALYSES

Two interim analyses will take place when 150 and 200 total subjects in the Per-Protocol Analysis Set 1 have completed 24 weeks follow-up or discontinue earlier. Each interim analysis will test both non-inferiority and futility of lenvatinib 14 mg arm $ORR_{24W}$ compared to lenvatinib 18 mg arm $ORR_{24W}$ based on investigator assessment. An O'Brien-Fleming stopping boundary (Lan and DeMets, 1983; DeMets and Ware, 1980) will be used for efficacy. An interpolated non-binding stopping boundary will be used for futility, which will spend $\beta$=0.005 and $\beta$=0.10 at the first and second interim analyses, respectively.

The stopping boundaries on the *P* value scale and the cumulative error probabilities spent at each interim analysis and final analysis are shown in Table below. For example, at the second interim analysis, non-inferiority in $ORR_{24W}$ will be claimed if the 1-sided *P* value is ≤0.014; futility will be claimed if the 1-sided *P* value is ≥0.207.

| Analysis # | Cumulative α Spent | Efficacy Boundary (*P* value) | Cumulative β Spent | Futility Boundary (*P* value) |
|---|---|---|---|---|
| Interim Analysis #1 | 0.005 | 0.005 | 0.005 | 0.776 |
| Interim Analysis #2 | 0.015 | 0.014 | 0.10 | 0.207 |
| Final Analysis | 0.05 | 0.045 | 0.2 | 0.045 |

Safety and efficacy monitoring will be performed by an independent DMC. The function and membership of the DMC are described in the DMC charter. The recommendations concerning continuation or termination of the trial, or regarding modification of the trial or informed consent documents will be reached by the DMC based on the review of study safety and efficacy data. Determination of whether lenvatinib 14 mg can be used as an alternative dosing strategy will be based on clinical judgment by the Sponsor in consultation with the independent DMC (as specified in the DMC charter) by assessing risks and benefits according to the totality of data at either of the interim or final analyses.

The interim analyses together with the safety monitoring DMC analyses will be performed by an independent statistical reporting team, such as a statistical contract research organization.

DMC meetings for the two protocol specified interim analyses have been carried out on 13 Mar 2019 and 12 Jun 2019. The DMC reviewed the interim analyses results and the study continues as planned.

# 7 CHANGES IN THE PLANNED ANALYSES

Not applicable.

# 8 DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

## 8.1 Visit Windows

Visit windows will be defined to be upper and lower bounds of 3 days of the scheduled visit, following the protocol, which states that efforts should be made to conduct study visits (and safety assessments) on the day scheduled (± 3 days).

Tumor assessments are to be performed every 8 weeks (during the 8$^{th}$ week, starting from the date of randomization), or sooner if clinically indicated, until progressive disease during Randomization Phase. If a subject discontinues from study treatment without disease progression, tumor assessments should continue to be performed every 8 weeks until documentation of disease progression or beginning a new anticancer treatment during Follow-up Period.

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 (eg, worst grade laboratory results and shift tables) will include all postbaseline assessments, including those scheduled and unscheduled visits.

## 8.2 Baseline Definitions

For safety assessments, the baseline value is defined as the last non-missing measurement taken on or prior to the first dose date. The first dose date is defined as the earliest date of non-zero dose administration of either study drug (ie, lenvatinib or everolimus).

Study day is defined as date of assessment – first dosing date + 1 for any assessment done on or after first dosing date; otherwise, study day is defined as date of assessment – first dosing date.

## 8.3 Imputation of Missing Data

Unless specified otherwise, no data imputation will be applied for missing safety and efficacy evaluations. For analysis and reporting purpose, partial dates for adverse events, prior and concomitant therapies, disease diagnosis date, and start date of new anticancer therapy will be imputed if needed. Partial dates for laboratory values, vital signs, and ECGs will not be imputed.

The imputation rules will be specified in study analysis dataset specification with more details.

## 8.4 Variable Derivations

### 8.4.1 Duration of Events Prior to Randomization

For the following disease characteristics and prior therapies, the duration in months will be calculated as the date of randomization minus the date of event (disease diagnosis, disease progression, prior therapies, etc.) and then divided by 30.4375.

The duration in years will be calculated as: (date of randomization – date of event)/365.25.

- Time since the first RCC diagnosis to date of randomization (months)
- Time since last disease progression to date of randomization (months)
- Time from end of last therapy to date of randomization (months)
- Time from prior surgery to date of randomization (months)
- Time from last radiotherapy to date of randomization (months)

## 8.5 Pharmacokinetics/Pharmacodynamics Data Handling

Details on calculating PK parameters, the way that the values of below limit of quantification (BLQ) to be treated, and the analyses to be conducted on PK/PD and exposure-response relationships will be specified in a separate plan.

When developing individual concentration-time profiles, BLQ values will be replaced with 0 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 to 0. If the proportion of values reported as BLQ is more than 50%, no summary statistics should be presented 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 9.4 or higher, and/or other validated statistical software as required.

# 11 MOCK TABLES, LISTINGS, AND GRAPHS

The study tables, listings, and graphs (TLGs) shells will be provided in a separate document, which will show the content and format of all the TLGs in details.

# 12 REFERENCES

Rothmann M, Li N, Chen G, Chi GYH, Temple RT, Tsou HH. Design and analysis of noninferiority mortality trials in oncology. Stat Med. 2003;22(2):239-64.

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(2):228-47.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika, 1983;70:659-63.

DeMets DL and Ware JH. Group sequential methods for clinical trials with a one-sided hypothesis. Biometrika. 1980;67:651-60.

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.

# 13 APPENDICES

## 13.1 Efficacy and Futility Boundaries at Each Interim Analysis and Final Analysis

There will be 2 interim analyses and one final analysis. The 2 interim analyses will be performed only in the Per-Protocol Analysis Set 1 (based on investigator assessment). The final analysis will be performed in the Per-Protocol Analysis Set 1 as the primary analysis set, and will also be performed in the Per-Protocol Analysis Set 2 and in the Full Analysis Set as secondary analysis sets.

The tables below list the treatment differences in $ORR_{24w}$ between lenvatinib 14 mg + everolimus arm and lenvatinib 18 mg + everolimus arm corresponding to the efficacy and futility boundaries at each interim analysis for different $ORR_{24w}$ in lenvatinib 18 mg + everolimus arm.

**First interim analysis, non-inferiority stopping boundary OR=1.788** (non-inferiority will be claimed if observed OR ≥1.788 at first interim analysis. Correspondingly, depending on $ORR_{24W}$, non-inferiority will be claimed if observed treatment difference between lenvatinib 14 mg arm and lenvatinib 18 mg arm is ≥δ).

| $ORR_{24W}$ in lenvatinib 18 mg arm | $\delta$ = Treatment difference (lenvatinib 14 mg arm - lenvatinib 18 mg arm) |
|---|---|
| 10% | 7% |
| 20% | 11% |
| 30% | 13% |
| 40% | 14% |
| 50% | 14% |
| 60% | 13% |

**First interim analysis, futility stopping boundary OR=0.590** (futility will be claimed if observed OR ≤0.590 at first interim analysis. Correspondingly, depending on $ORR_{24W}$, futility will be claimed if the observed treatment difference between lenvatinib 14 mg arm and lenvatinib 18 mg arm is ≤δ).

| $ORR_{24W}$ in lenvatinib 18 mg arm | $\delta$ = Treatment difference (lenvatinib 14 mg arm - lenvatinib 18 mg arm) |
|---|---|
| 10% | -4% |
| 20% | -7% |
| 30% | -10% |
| 40% | -12% |
| 50% | -13% |
| 60% | -13% |

**Second interim analysis, non-inferiority stopping boundary OR=1.436** (non-inferiority will be claimed if the observed OR ≥1.436 at second interim analysis. Correspondingly, depending on $ORR_{24W}$, non-inferiority will be claimed if the observed treatment difference between lenvatinib 14 mg arm and lenvatinib 8 mg arm is ≥δ).

| $ORR_{24W}$ in lenvatinib 18 mg arm | $\delta$ = Treatment difference (lenvatinib 14 mg arm - lenvatinib 18 mg arm) |
|---|---|
| 10% | 4% |
| 20% | 6% |
| 30% | 8% |
| 40% | 9% |
| 50% | 9% |
| 60% | 8% |

**Second interim analysis, futility stopping boundary OR=0.962** (futility will be claimed if the observed OR ≤0.962 at the second interim analysis. Correspondingly, depending on $ORR_{24W}$, futility will be claimed if the observed treatment difference between lenvatinib 14 mg arm and lenvatinib 18 mg arm is ≤δ).

| $ORR_{24W}$ in lenvatinib 18 mg arm | $\delta$ = Treatment difference (lenvatinib 14 mg arm - lenvatinib 18 mg arm) |
|---|---|
| 10% | -0.3% |
| 20% | -0.6% |
| 30% | -0.8% |
| 40% | -0.9% |
| 50% | -1.0% |
| 60% | -0.9% |

**Final analysis, OR=1.128** (non-inferiority will be claimed if observed OR≥1.128 at the final analysis. Correspondingly, depending on $ORR_{24W}$, non-inferiority will be claimed if the observed treatment difference between lenvatinib 14 mg arm and lenvatinib 18 mg arm is ≥δ).

| $ORR_{24W}$ in lenvatinib 18 mg arm | $\delta$ = Treatment difference (lenvatinib 14 mg arm - lenvatinib 18 mg arm) |
|---|---|
| 10% | 1% |
| 20% | 2% |
| 30% | 3% |
| 40% | 3% |
| 50% | 3% |
| 60% | 3% |

The above boundaries are based on the plan that the two interim analyses will take place when the total number of 150 subjects and 200 subjects in the Per-Protocol Analysis set 1 have completed 24 weeks of follow-up or discontinued early.

Safety and efficacy monitoring will be performed by an independent DMC. The function and membership of the DMC are described in the DMC charter. The recommendations concerning continuation or termination of the trial, or regarding modification of the trial or informed consent documents will be reached by the DMC based on the review of study safety and efficacy data. Determination of whether lenvatinib 14 mg can be used as an alternative dosing strategy will be based on clinical judgment by the Sponsor in consultation with the independent DMC (as specified in the DMC charter) by assessing risks and benefits according to the totality of data at either of the interim or final analyses.

The interim analyses together with the safety monitoring DMC analyses will be performed by an independent statistical reporting team, such as a statistical contract research organization.

# SIGNATURE PAGE

Author:

PPD

PPD

Date

Oncology Business Group, Eisai Inc.

Approval:

PPD

Date

Oncology Business Group, Eisai Inc.

PPD

Date

Oncology Business Group, Eisai Ltd.