STATISTICAL ANALYSIS PLAN

Study Protocol Number:	E7080-G000-211 (2-Arm Design)
Study Protocol Title:	A Multicenter, Randomized, Double-Blind Phase 2 Trial of Lenvatinib (E7080) in Subjects With ^{131}I-Refractory Differentiated Thyroid Cancer to Evaluate Whether an Oral Starting Dose of 18 mg Daily Will Provide Comparable Efficacy to a 24-mg Starting Dose, But Have a Better Safety Profile
Date:	20 March 2020
Version:	Final Version 1.0

1 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

2 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
^{131}I	Radioiodine
β-hCG	beta-human chorionic gonadotropin
AE	adverse event
AJCC	American Joint Committee on Cancer
ALT	alanine aminotransferase
anti-Tg	antithyroglobulin autoantibodies
AST	aspartate aminotransferase
ATC	anatomical therapeutic class
BMI	body mass index
BOR	best overall response
BP	blood pressure
BUN	blood urea nitrogen
CBR	clincial benefit rate
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CPK	creatine phosphokinase
CR	complete response
CRF	case report form
CSE	clincially significant treatment-emergent adverse event
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CS	clinically significant
DCR	disease control rate
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
FDA	US Food and Drug Administration
HRQoL	Health-Related Quality of Life

Abbreviation	Term
INR	International Normalized Ratio
KM	Kaplan-Meier
LDH	lactate dehydrogenase
LLT	lower level term
LVEF	left ventricular ejection fraction
MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
MUGA	multiple-gated acquisition
NCS	not clinically significant
NYHA	New York Heart Association
ORR	objective response rate
OS	overall survival
PD	progression disease
PFS	progression-free survival
PFS2	progression-free survival after next line of treatment
PK	pharmacokinetic
PR	partial response
PS	performance status
PT	preferred term
pTNM	primary tumor-node-metastasis staging
Q1	1st Quartile
Q3	3rd Quartile
QD	once daily
RBC	red blood cells
RECIST 1.1	Response Evaluation Criteria in Solid Tumors Version 1.1
RR-DTC	radioiodine-refractory differentiated thyroid cancer
RR	respiratory rate

Abbreviation	Term
SAE	treatment-emergent serious adverse event
SAP	statistical analysis plan
SD	stable disease
SI	Système International
SOC	system organ class
T4	thyroxine
TEAE	treatment-emergent adverse event
TLG	tables, listings, and graphs
Tg	thyroglobulin
TSH	thyroid-stimulating hormone
UPCR	urine protein-to-creatinine ratio
VEGF	vascular endothelial growth factor
WBC	white blood cells
WHO DD	World Health Organization Drug Dictionary

3 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report results for Eisai Protocol E7080-G000-211, *A Multicenter, Randomized, Double-Blind Phase 2 Trial of Lenvatinib (E7080) in Subjects With Radioiodine-Refractory Differentiated Thyroid Cancer (RR-DTC) to Evaluate Whether an Oral Starting Dose of 18 mg Daily Will Provide Comparable Efficacy to a 24-mg Starting Dose, But Have a Better Safety Profile*. This SAP focuses on the primary analysis planned and specified in Protocol Amendment 05 (dated as 21 May 2019). This SAP will be finalized and approved before database lock.

This study was initially designed as a 3-arm trial to evaluate whether an oral lenvatinib starting dose of 20 mg or 14 mg daily will provide comparable efficacy to a 24-mg lenvatinib starting dose, but have a better safety profile in subjects with ^{131}I-refractory differentiated thyroid cancer (DTC). Results of new modeling and simulation tests became available, indicating that the planned alternate starting doses might not be the appropriate doses to evaluate the study hypothesis. Agreement was then reached with the FDA and EMEA to redesign the study to a 2-arm design, with 1 proposed alternate dose.

As of Protocol Amendment 03, the study design was changed to a 2-arm design to evaluate whether an oral lenvatinib starting dose of 18 mg daily will provide comparable efficacy to a 24-mg lenvatinib starting dose, but have a better safety profile. Consequentially, the sample size was changed to 152 subjects, with 76 subjects per arm.

There will be 2 separate Clinical Study Reports (CSRs): one for the 3-arm design with 41 randomized subjects and one for the 2-arm design with 152 randomized subjects. The purpose of this SAP is to describe the procedures and the statistical methods that will be used to analyze and report results for Eisai Protocol E7080-G000-211, 2-arm design. This SAP does not cover the analyses of pharmacokinetic (PK)/pharmacodynamics, biomarkers, and HRQoL. The analyses of PK/pharmacodynamics data from both parts will be analyzed and reported separately. The Health-Related Quality of Life (HRQoL) will also be reported separately. However, results for both analyses will be briefly summarized in the CSR for the 2-arm design.

3.1 Study Objectives

3.1.1 Primary Objective

To determine whether a starting dose of lenvatinib 18 mg once daily (QD) will provide comparable efficacy (based on objective response rate [ORR] at 24 weeks [ORR_{24wk}]) with an improved safety profile compared to 24 mg QD (based on treatment-emergent adverse events [TEAEs] of Grade 3 or higher in the first 24 weeks after randomization).

3.1.2 Secondary Objectives

- To evaluate progression-free survival (PFS) in subjects treated with lenvatinib doses of 24 mg and 18 mg QD
- To evaluate the PFS after next line of treatment (PFS2) in subjects treated with lenvatinib doses of 24 mg and 18 mg QD.
- To evaluate the safety and tolerability of lenvatinib doses of 24 mg and 18 mg QD.
- To evaluate the PK/pharmacodynamic relationship between exposure and biomarkers/efficacy/safety, using a mechanistically based approach, if possible.
- To evaluate the impact of lenvatinib treatment on HRQoL as measured by the instruments EQ-5D-3L and FACT-G.

3.1.3 Exploratory Objectives

- To explore overall survival (OS) in subjects treated with lenvatinib doses of 24 mg and 18 mg QD.
- To explore thyroglobulin, thyroid-stimulating hormone (TSH), and other serum biomarkers as potential biomarkers for tumor response.
- To explore DNA sequence variants in genes that might influence PK, safety, or pharmacodynamics data.

3.2 Overall Study Design and Plan

This was a multicenter, randomized, double-blind study being conducted as a postmarketing commitment to the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) to evaluate whether there was a lower starting dosage of lenvatinib other than 24 mg QD that could provide comparable efficacy but have a better safety profile in subjects with RR-DTC with radiographic evidence of disease progression (PD) within the prior 12 months.

Eligible subjects should have measurable disease according to Response Evaluation Criteria in Solid Tumors Version 1.1 (RECIST 1.1) and evidence of PD within 13 months before signing informed consent (both assessed by independent radiologic review) and were randomly assigned to treatment in a 1:1 ratio to receive lenvatinib 24 mg or 18 mg orally QD. The total sample size is 152 subjects (76 subjects per arm). Randomization was stratified by age (≤65 years or >65 years) and Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0 vs 1 or 2. Study ended when the last subject enrolled completed the Week 24 tumor assessments or discontinued study drug before Week 24. Subjects received study treatment until disease progression, development of unacceptable toxicity, subject requested to discontinue, withdrew consent or was lost to follow-up, end of the study, or until study termination by the sponsor. After PD was observed, subjects were followed for PFS2 and survival.

This study consisted of 2 phases: the Prerandomization/Pretreatment Phase and the Randomization/Treatment Phase (see Figure 1). The Schedules of Procedures and Assessments for both study phases are explained in Table 7 of Protocol Amendment 05. Per protocol, tumor assessments were performed during the Prerandomization Phase and then every 8 weeks (within the 8th week) from the date of randomization during treatment cycles in the Randomization Phase. Subjects who discontinued treatment without PD in the Randomization Phase continued to undergo tumor assessments every 8 weeks until PD was documented or another anticancer therapy was initiated. Subjects were then followed every 12 weeks (±1 week) for survival, PFS2 and all subsequent anticancer treatments received (unless this information was not allowed to be provided due to confidentiality) had been recorded until the End of Study.

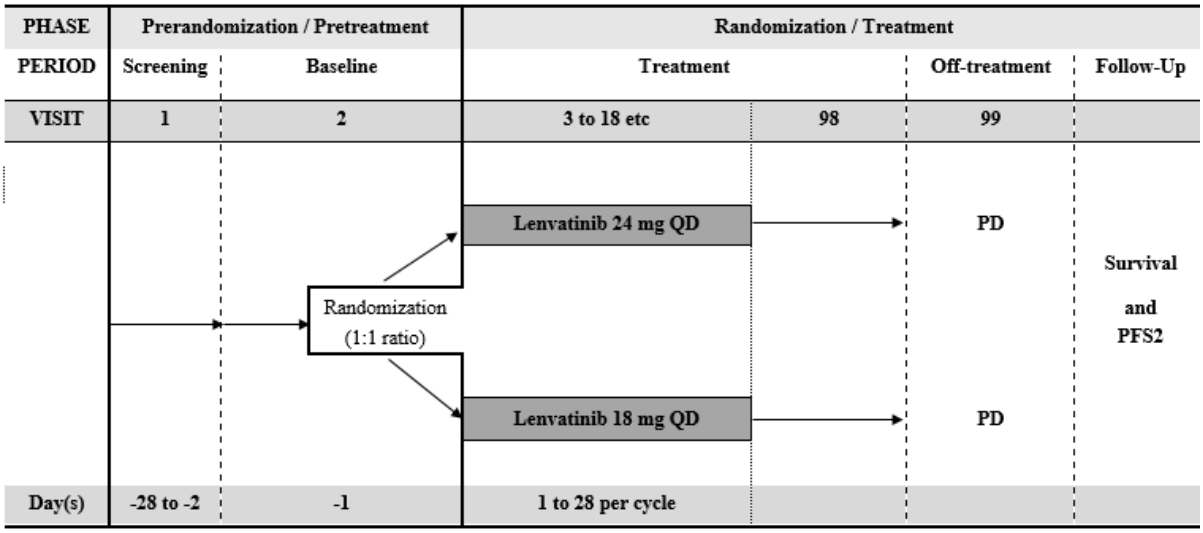

Figure 1. Study Design for Study E7080-G000-211 (as of Protocol Amendment 05)

PD = progressive disease, PFS2 = progression-free survival after next line of anticancer treatment, QD = once daily.

4 DETERMINATION OF SAMPLE SIZE

Sample size determination was based on the number of subjects required to detect noninferiority of the primary efficacy endpoint, ORR_{24wk}, comparing the 18 mg arm to the 24 mg arm. Based on the data from Study 303, the ORR_{24wk} for the lenvatinib 24 mg arm was 54.4% (142 responders, N=261) and 0.8 % for the placebo arm (1 responder, N=131) and the lower 95% CI of the Odds Ratio for lenvatinib 24 mg vs placebo was 21.4. Assuming a 70% retention of the effect of lenvatinib 24 mg versus placebo, the noninferiority margin on the Odds Ratio scale was estimated to be 0.40 (i.e, Ha: Odds Ratio [18 mg/24 mg] >0.4).

A sample size of 152 subjects (76 per arm) would provide a statistical power of 80% to declare noninferiority, assuming a 1-sided alpha of 0.025, an ORR_{24wk} of 54.4% for lenvatinib 24 mg and a true Odds Ratio of 1, with a noninferiority margin of 0.4.

For the primary safety endpoint, based on data from Study 303, where 70.5% of subjects in the lenvatinib 24 mg arm had a TEAE of Common Terminology Criteria for Adverse Events (CTCAE) Grade 3 or higher during the first 24 weeks, the sample size of 152 subjects provides a precision for the observed differences between the arms with half-widths of the 95% CI of about 15%.

5 STATISTICAL METHODS

All descriptive statistics for continuous variables will be reported using mean, standard deviation, median, 1st quartile (Q1), 3rd quartile (Q3), and minimum and maximum values and will be specified in the outputs. Categorical variables will be summarized as the number (percentage) of subjects.

Unless otherwise specified, summary reporting will be presented by lenvatinib starting dose (24 mg, 18 mg) and overall treatment; the analysis set to be used is explained in Table1 (Section 5.2.1).

5.1 Study Endpoints

All study planned analysis endpoints are explained in this section. Per protocol, tumor assessment related endpoints will be defined based on investigator review per RECIST 1.1. The tumor assessment schedules are described in Table 7 of Protocol Amendment 05.

5.1.1 Primary Endpoints

- ORR_{24wk} as assessed by the investigator using RECIST 1.1. ORR_{24wk} is defined as the proportion of subjects with best overall response (BOR) of complete response (CR) or partial response (PR) at the Week 24 time point or earlier. The BOR for each subject will be determined based on RECIST 1.1 without a requirement for response confirmation.
- Rate of TEAEs with CTCAE grades of 3 or higher within 24 weeks after randomization (as of the Week 24 time point).

5.1.2 Secondary Endpoints

- 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
- PFS2, defined as the time from randomization to second objective disease progression (occurring during treatment with next line of anticancer therapy), or death from any cause, whichever occurs first
- Overall safety profile and tolerability
- Time to treatment discontinuation due to an adverse event (AE)
- Number of dose reductions

- Time to first dose reduction
- *Plasma PK lenvatinib exposure parameters
- *Interrelationships of lenvatinib exposure, changes in thyroglobulin, TSH, or other exploratory serum biomarkers, and changes in tumor burden and PFS
- *Relationship of lenvatinib exposure and changes in blood pressure (BP), and AEs of weight loss, fatigue, nausea, vomiting, diarrhea, and proteinuria CTCAE grades derived from urine protein measurements
- *Impact of lenvatinib treatment on HRQoL as assessed using the validated instruments EQ-5D-3L and FACT-G

* Will be provided in separate reports.

5.1.3 Exploratory Endpoints

- Duration of response, defined as the time from the initial achievement of a response (CR or PR) to the date of first documentation of disease progression, or the date of death, whichever occurs first
- Disease control rate (DCR), defined as the proportion of subjects who have BOR of CR, PR, or stable disease (SD). BOR of SD must be achieved at least 7 weeks after randomization
- Clinical benefit rate (CBR), defined as the proportion of subjects who have BOR of CR, PR, or durable SD (duration of SD ≥23 weeks after randomization)
- OS, measured from the date of randomization until date of death from any cause
- *Associations between objective tumor response and serum thyroglobulin (accounting for anti-Tg), TSH, and other serum biomarkers (including VEGF, Ang-2, sTie-2, and FGF23)
- *Association between any observed DNA sequence variability and PK, pharmacodynamic, and clinical outcome measures including efficacy and safety-related endpoints

* Will be provided in a separate report.

5.2 Study Subjects

5.2.1 Definitions of Analysis Sets

Full Analysis Set will include all randomized subjects. This will be the analysis set for all efficacy evaluations, which will be analyzed according to the treatment randomized, regardless of the treatment actually received.

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

Pharmacokinetic (PK) Analysis Set will include all subjects who received at least one dose of study drug and who have evaluable lenvatinib plasma concentration data.

The parameters analyzed for each analysis set are listed in Table 1.

Table 1 Analysis Sets

Analyses to be performed	Analysis Sets			
	Enrolled Subjects	Full Analysis Set	Safety Analysis Set	PK Analysis Set
Protocol Deviations		●		
Disposition	●			
Demography & Baseline Characteristics		●		
Disease History		●		
Prior & Concomitant Medications		●		
Progression-Free Survival		●		
Progression-Free Survival 2		●		
Overall Survival		●		
Tumor Response		●		
Drug Exposure			●	
Adverse Events			●	
Deaths	●		●	
Laboratory Tests			●	
Vital Signs			●	
Electrocardiograms (ECGs)			●	
Echocardiography			●	
Pharmacokinetics				●
Clinical Biomarkers			●	

5.2.2 Subject Disposition

The number of subjects enrolled and the number (percentage) of subjects who were screen failures and the primary reason for screen failure will be summarized based on data from the Case Report Form (CRF) of Screening Disposition.

The number of subjects screened, randomized, treated, and with major protocol deviations will be counted by site. Subject disposition on study treatment (treatment ongoing at data cutoff date, treatment discontinued, and the primary reason for discontinuation) will be

summarized. Study status at data cutoff and reason for withdrawal from the study will also be summarized using data from survival follow-up.

5.2.3 Protocol Deviations

Major protocol deviation criteria will be established and all subjects with potential major protocol deviations will be identified and reviewed in a blinded manner prior to database lock. Major protocol deviations will be summarized by each category and will be listed.

5.2.4 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized. Continuous demographic and baseline variables include age, body weight and height, body mass index (BMI), categorical variables include sex, age group (≤65, >65), race, ethnicity, region, body weight group (<60 kg, ≥60 kg), ECOG (0, 1, 2), New York Heart Association (NYHA) classification, and baseline TSH (≤0.5, >0.5 to 2.0, >2.0 to 5.5 μIU/mL).

DISEASE HISTORY

Disease history and characteristics at study entry will be summarized based on data recorded on the CRFs, including thyroid cancer pTNM staging at diagnosis, DTC subtype diagnosis, stage group at diagnosis, time since diagnosis (original and metastatic), and prior therapy (anticancer medications, radiotherapy, previous radioiodine therapy). In addition, locally advanced or metastatic DTC will be characterized using the location of both target and nontarget lesions at screening, and summarized. The number of metastatic sites per subject will also be included in the summary table.

MEDICAL HISTORY

Subjects with any previous medical condition and current medical condition, as recorded on the CRF, will be summarized by system organ class (SOC) and preferred term (PT). The Medical Dictionary for Regulatory Activities (MedDRA) version 23.0 will be used.

5.2.5 Prior and Concomitant Therapy

All investigator terms for medications recorded in the CRF will be coded to an 11-digit code using the World Health Organization Drug Dictionary (WHO DD) Version (WHODDMAR20B3). Prior medications will be defined as medications that started before the first dose of study drug. Concomitant medications will be defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 28 days after the subject's last dose. Medications that cannot be determined to be prior/concomitant/posttreatment because of missing or incomplete dates will be regarded as concomitant.

The number (percentage) of subjects who took prior medications will be summarized by anatomical class (Anatomical Therapeutic Chemical [ATC] Level 1), pharmacological class (ATC Level 3), and World Health Organization Drug Name (preferred term). The same summary will be repeated for concomitant medications except for thyroxine suppression therapy and antihypertensive therapy. Concomitant thyroxine suppression therapy and antihypertensive therapy will be summarized separately. In addition, there will be a separate summary for P-glycoprotein inhibitors and/or inducers taken as concomitant medications.

5.2.5.1 Previous Therapies

Previous chemotherapy, VEGF-targeted therapy, radiotherapy, radioiodine therapy, and antithyroid cancer surgery will be summarized, and listings will be provided for previous anticancer medications, radioiodine therapy, and radiotherapy.

Prior database lock, previous chemotherapy and previous VEGF-targeted therapy will be identified by the clinical and coding team based on data from the Previous Anticancer Medications CRF page. Similarly, previous antithyroid cancer surgery will be identified from the Medical History CRF page.

5.2.6 Treatment Compliance

Treatment compliance will be summarized by the number (percentage) of subjects with any drug interruption not due to investigator request. In addition, the percentage of days that study drug was interrupted not due to investigator request will be calculated.

5.3 Data Analysis General Considerations

5.3.1 Pooling of Centers

This is a multicenter study. Subjects from all centers will be pooled for all analyses.

5.3.2 Adjustments for Covariates

Not applicable.

5.3.3 Multiple Comparisons/Multiplicity

Not applicable.

5.3.4 Examination of Subgroups

No subgroup analyses are planned for this study.

5.3.5 Handling of Missing Data, Dropouts, and Outliers

Rules for handling full/partial missing dates, and the censoring rules for PFS, PFS2, and OS are explained in Section 8.4. The handling of missing data in the analyses of ORR, CBR and

DCR is explained in Sections 5.4. Any potential outlier values will be investigated and will be analyzed in the locked database as originally reported.

5.3.6 Other Considerations

No other consideration.

5.4 Efficacy Analyses

5.4.1 Primary Efficacy Analyses

The noninferiority analysis of ORR_{24wk} between lenvatinib starting doses of 24 mg and 18 mg will be performed at the 1-sided alpha of 0.025, based on the calculated odds ratio of ORR_{24wk} response (18 mg vs 24 mg) along with its 95% confidence interval (CI) using the Cochran-Mantel-Haenszel (CMH) method, stratified by the randomization stratification factors. The test will be performed per the 95% CI using the noninferiority margin of 0.4. Noninferiority will be declared if the lower limit of the 95% CI for the odds ratio is greater than 0.4.

The point estimate of ORR_{24w} for each group will be summarized with the corresponding 95% CI using asymptotic normal approximation. Treatment difference (18 mg minus 24 mg) will also be estimated with its 95% CI using asymptotic normal approximation.

SENSITIVITY ANALYSIS

A sensitivity analysis will be performed for overall ORR using the same methods as used for ORR_{24wk}.

5.4.2 Secondary Efficacy Analyses

- The following analyses will be performed on PFS, and censoring rules are detailed in Section 8.4:
  - The difference in PFS between lenvatinib starting dose of 24 mg and 18 mg using stratified log-rank test with the randomization stratification factors, tested at an alpha level of 0.05
  - The estimate of the hazard ratio calculated from Cox proportional hazard model on PFS with lenvatinib starting dose as the independent variable and stratified by the randomization stratification factors with a 2-sided 95% CI
  - Kaplan-Meier (KM) estimated quartiles of PFS and their 95% CIs (using the generalized Brookmeyer and Crowley method) per lenvatinib starting dose
  - KM estimated PFS rates at 6, 12, 18 and 24 months and their 95% CIs (using Greenwood log-log formula) per lenvatinib starting dose
  - KM plot of estimated PFS over time

- The analyses of PFS will be repeated for PFS2, and the censoring rule is detailed in Section 8.4

5.4.3 Other Efficacy Analyses

Duration of response (months), defined as the time from the initial achievement of a response (CR or PR) to the date of first documentation of PD or the date of death, whichever occurs first. It will be summarized for subjects with BOR of CR or PR.

For both endpoints, the censoring rules are the same as those for PFS censoring, when applicable. KM estimated quartiles and their 95% CIs per lenvatinib starting dose, using the generalized Brookmeyer and Crowley method, will be provided.

DCR and CBR will be evaluated using the same method as that used for ORR_{wk24}.

OS will be evaluated using the same method as that used for PFS, and censoring rule is detailed in Section 8.4.

In addition, waterfall plots will be presented by lenvatinib starting dose for the percentage changes from Baseline to postbaseline nadir value in the sum of diameters of target lesions.

5.5 Pharmacokinetic, Pharmacodynamic, Pharmacogenomic, and Other Biomarker Analyses

5.5.1 Pharmacokinetic Analyses

Individual listings with a summary of plasma concentrations of lenvatinib and a figure containing lenvatinib plasma concentrations versus PK sampling time will be provided by subject and lenvatinib starting dose (18 vs 24 mg) for subjects in the Pharmacokinetic Analysis Set. Other PK analyses will be described in a separate SAP for PK/PD/Biomarkers.

5.5.2 Pharmacodynamic, Pharmacogenomic, and Other Biomarker Analyses

Summary of TSH, thyroglobulin (Tg), anti-Tg antibodies and other serum biomarkers (including VEGF, Ang-2, sTie-2, FGF23) will be provided by visit using Safety Analysis Set. Listings and Box and Whisker plots of serum pharmacodynamic results over time and the figure for the percentage change from Baseline for serum pharmacodynamic results over time will be provided.

Other pharmacodynamic and biomarker analyses will be described in a separate SAP for PK/PD/Biomarker.

5.6 Safety Analyses

All safety analyses will be performed on the Safety Analysis Set. Safety data, presented by the starting lenvatinib dose of 24 mg or 18 mg and overall will be summarized on an "as treated" basis using descriptive statistics (i.e., n, mean, standard deviation, median, minimum, and maximum values for continuous variables; n [%] for categorical variables). Safety variables include TEAEs, TEAEs with CTCAE Grade ≥3, intolerable TEAEs of CTCAE Grade 2 (based on data collected in CRF), treatment-related TEAEs, treatment-emergent serious adverse events (SAEs), clinical laboratory parameters, vital signs, 12-lead ECG results, ECOG PS scores, and left ventricular ejection fraction (LVEF). Abnormal values will be flagged.

Study Day 1 for all safety analyses is defined as the date of the first dose of study drug.

5.6.1 Extent of Exposure

Treatment period consisted of 28-day blinded study treatment cycles (i.e., 1 to 28 days per cycle) (Section 3.2). Per protocol, subjects received 1 of the 2 starting doses, lenvatinib 24 mg/day or lenvatinib 18 mg/day. Dose reductions would occur in succession based on the subject's previous dose level (24, 20, 14, 10, and 8 mg QD, or 18, 14, 10, 8, and 4 mg QD, respectively).

The following analyses will be performed:

- Duration of Treatment (months): (last dose date – first dose date + 1) / (365.25/12)
- Number of Subject-months: sum of treatment time (in months) for all subjects (including dose interruption).
- Total dose (mg) taken per subject
- Dose intensity (mg/day) per subject (i.e., total dose/treatment duration), relative dose intensity (%) per subject (i.e., percentage of dose intensity relative to the planned dose at Baseline)
- Dose modification (reduction, interruption), frequency of dose modification
- Treatment discontinuation due to AE (from treatment disposition page)
- Time to first dose reduction (weeks), time to treatment discontinuation due to AE (weeks). Both will be analyzed using KM method (same as for PFS, see Section 5.4.1) and estimated quartiles and the 95% CI will be presented.

5.6.2 Adverse Events

The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the MedDRA. Adverse events will be coded to MedDRA (version 23.0) lower level term (LLT) closest to the verbatim term. The linked

MedDRA PT and primary SOC will also be captured in the database. The CTCAE grading (version 4.03) will be included in the database.

A TEAE is defined as an AE that emerged during treatment (from time of first dose up to 28 days following the last dose of study drug), having been absent pretreatment (Baseline) or

- Reemerged during treatment, having been present pretreatment (Baseline) but stopped before treatment, or
- Worsened in severity during treatment relative to the pretreatment state, when the AE was continuous

Only TEAEs will be summarized. A subject will be counted only once within each category (i.e., SOC, PT, worst grade, or other specific category to be summarized in table) even if the subject experienced more than 1 TEAE within the category. All AEs, treatment-emergent or otherwise, will be presented in subject data listings.

For the analysis of the primary safety endpoint of rate of TEAEs with CTCAE Grade ≥3 within 24 weeks after randomization (as of the Week 24 time point), the number (percentage) of subjects with TEAEs of CTCAE Grade ≥3 as of Week 24 will be reported; the difference in frequency between the treatments (18 mg minus 24 mg) will be presented with 95% CI, using asymptotic normal approximation. In addition, the number (percentage) of subjects with TEAEs with CTCAE Grade >=3 as of the Week 24 will be summarized by SOC and PT.

The incidence of TEAEs will be reported as the number (percentage) of subjects with TEAEs by SOC, PT and worst CTCAE grade. The number (percentage) of subjects with TEAEs will also be summarized by PT and worst CTCAE grade.

The number (percentage) of subjects with treatment-related TEAEs, TEAEs leading to treatment discontinuation, TEAEs leading to dose reduction/interruption, will be summarized by SOC, PT and worst CTCAE grade. All SAEs, treatment-related SAEs, fatal TEAEs, and nonfatal SAEs will be summarized by SOC and PT. Treatment-related TEAEs include those events considered by the investigator to be related to study treatment or TEAEs with a missing causality on the CRF.

In addition, summaries of clinically significant TEAEs (CSEs) (i.e., subject incidence, time to first onset of Grade 3 or 4) will be provided. Appendix 13.2 includes the list of CSEs. These CSEs include those AEs known to be of concern with lenvatinib or of VEGF-targeted agents in general.

TEAEs, SAEs, treatment-related TEAEs, treatment-related SAEs, CSEs will also be summarized by PT in decreasing frequency.

The number (percentage) of subjects with all deaths and deaths on treatment or within 30 days of last lenvatinib dose will be summarized.

Subject data listings for (1) all AEs, (2) AEs of CTCAE Grade 2 (intolerable) or above, (3) serious adverse events, (4) fatal AEs, (5) all deaths, (6) AEs leading to treatment discontinuation, and (7) AEs leading to dose reduction or interruption will be provided.

5.6.3 Laboratory Values

Clinical laboratory test parameters, per category, are listed in Table 2. The results will be standardized using System International (SI) units and categorized according to CTCAE grade (v4.03) (see protocol Appendix 13.8).

Table 2 Clinical Laboratory Tests

Category	Parameters
Hematology	Hematocrit, hemoglobin, MCH, MCHC, MCV, platelets, RBC count, and WBC count with differential (bands[a], basophils, eosinophils, lymphocytes, monocytes, neutrophils), INR[b]
Chemistry	Bicarbonate[a], chloride, potassium, sodium, BUN or urea, creatinine, glucose, magnesium, phosphorus, calcium, albumin, total protein, alkaline phosphatase, ALT, AST, conjugated (direct) bilirubin, total bilirubin, TSH, Amylase[c], lipase, CPK[c], LDH[f], Total cholesterol, triglycerides, free T4 levels[f]
Urine Dipstick Testing[d]	glucose, hemoglobin (or blood), ketones, pH, protein[e], specific gravity

ALT = alanine aminotransferase, AST = aspartate aminotransferase, BUN = blood urea nitrogen, CPK = creatine phosphokinase, β-hCG = beta-human chorionic gonadotropin, INR = International Normalized Ratio, LDH = lactate dehydrogenase, MCH = mean corpuscular hemoglobin, MCHC = mean corpuscular hemoglobin concentration, MCV = mean corpuscular volume, RBC = red blood cells, T4 = thyroxine, TSH = thyroid stimulating hormone, UPCR = urine protein-to-creatinine ratio, WBC = white blood cells.

a: Optional if results cannot be obtained from the local laboratory.

b: INR should only be performed as part of the screening assessment and when clinically indicated.

c: Amylase isoenzymes (pancreatic and salivary type) and CPK isoenzymes (CK-MM and CK-MB) should be evaluated if amylase or CPK is greater than 3 × the upper limit of normal.

d: If urine dipstick testing suggests a urinary tract infection, or if clinically indicated, a urine microscopy, culture, and sensitivity should be performed at the institution's laboratory.

e: If urine protein is ≥2+, then a 24-hour urine collection or an immediate spot UPCR test should be done to quantify the 24-hour urine protein excretion.

f: For screening only.

Each parameter will be summarized or plotted by visit (with at least 10% of subjects) for actual value observed and change from Baseline. For thyroid stimulating hormone (TSH), the lowest postbaseline values will be summarized per category (≤0.5; >0.5–2.0; >2.0–5.5; >5.5 mIU/L).

Shift tables from Baseline to worst postbaseline CTCAE grade will be provided for hematology, chemistry, dipstick, and 24-hour urine collection. Grade 3 or 4 treatment-emergent laboratory results for hematology and chemistry will also be summarized.

5.6.4 Vital Signs

Each parameter of vital signs assessments (diastolic and systolic BP, pulse rate, respiratory rate, temperature) and body weight will be summarized/plotted by visit (only those having results for ≥10% of subjects) for the actual value observed and its change from Baseline. Hypertension, based on diastolic and systolic BP measurements, will also be summarized as shift from Baseline to worst postbaseline grade (see Table 3 for CTCAE grading).

Table 3 Blood Pressure Grading per CTCAE Version 4.03

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

5.6.5 Electrocardiograms

Each parameter of ECG assessments (heart rate, QT interval, QTcB interval, QTcF interval, RR interval) will be summarized by visit based on actual value observed and its change from Baseline.

ECG findings (normal; abnormal, not clinically significant [NCS]; abnormal, clinically significant [CS]) will also be summarized as shift from Baseline to the worst value during the treatment period. In addition, the number (percentage) of subjects with at least 1 postbaseline abnormal ECG result in QTc, using both Bazett's (QTcB) and Fridericia's (QTcF) correction formulas, during the treatment period will be summarized. Clinically abnormal ECG results in QTcB and QTcF will be categorized as follows:

- Absolute QTc interval prolongation:
  - ≤450 ms
  - >450 ms - ≤480 ms
  - >480 ms - ≤500 ms
  - >500 ms
- Increase from Baseline in QTc interval:
  - ≤30 ms
  - >30 ms - ≤60
  - >60 ms

5.6.6 Other Safety Analyses

Shift tables for ECOG PS from Baseline to worst postbaseline score (0 to 4) will be presented.

Descriptive statistics for LVEF as assessed by either echocardiogram or MUGA and changes from Baseline will be summarized.

5.7 Other Analyses

Health-related Quality of Life will be described in a separate SAP.

5.8 Exploratory Analyses

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

6 INTERIM ANALYSES

No interim analysis is planned for this study.

7 CHANGES IN THE PLANNED ANALYSES

This SAP focuses on the analyses for the 2-arm design of the study.

The Safety Analysis Set will be used for the summary table of biomarker parameters, while the Pharmacodynamic Analysis Set will be defined in a separate SAP for PK/PD analyses.

The exploratory endpoint of Duration of Clinical Benefit will not be analyzed in this study as it is not a widely used and well-defined endpoint; furthermore, it was not included in the RR-DTC Phase 3 study, E7080-G000-303.

Clinically abnormal ECG results based on QTcB and QTcF will be summarized the same way as done in Study E7080-G000-303. The categories are slightly different than the ones stated in the study protocol.

8 DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

8.1 Baseline

Baseline is defined as the nonmissing value most recently collected prior to the first dose of study drug.

8.2 Study Day

Study Day is defined as the day counting from the first lenvatinib dose date as follows:

- For a date on or after the first dosing date, Study Day= event date – first dose date of study drug +1
- For a date prior to the first dosing date, Study Day= event date – first dose date of study drug

8.3 Conversion Rules for Day to Month to Year

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

8.4 Censoring Rules

The censoring rules for PFS are explained in Table 4, based upon investigator's tumor assessments and the FDA Guidance for Industry, *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics* (2007).

For PFS2, if a subject did not have PD (solely based on the investigator-provided information on the Survival Status eCRF) after starting a new anticancer treatment, or did not die during the study, the subject will be censored on the date last known to be alive.

For OS, in the absence of confirmation of death, subjects will be censored either at the date last known to be alive or the date of data cutoff, whichever is earlier.

Table 4 Censoring Rules for Progression-Free Survival

No.	Situation	Date of Progression or Censoring	Outcome
1	No baseline or postbaseline tumor assessments	Date of randomization	Censored
2	Progression documented between scheduled visits	Date of first radiologic PD assessment	Progressed
3	No progression at the time of data cutoff	Date of last adequate radiologic assessment prior to or on date of data cutoff	Censored
4	New anticancer treatment started	Date of last adequate radiologic assessment prior to or on date of new anticancer treatment	Censored
5	Death before first PD assessment	Date of death	Progressed
6	Death between adequate assessment visits*	Date of death	Progressed
7	Death or progression after more than one missed visit** or after 28 days from the last dose of study treatment	Date of last adequate radiologic assessment before missed tumor assessments	Censored
8	Treatment discontinuation for reasons other than PD	Date of last radiologic assessment before treatment discontinuation	Censored

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

* Adequate tumor assessment is a radiologic assessment of CR, PR, SD, non-CR/non-PD or PD as determined by investigators.

** More than 1 missed visits is defined as having either 1 of the following 2 durations being longer than 18 weeks - 1 day, which is 125 days (= ((8+1) × 2 × 7) – 1) for subjects on the every-8-week tumor assessment schedule in this study:

a. Duration between 2 consecutive tumor assessments

b. Duration between the last adequate tumor assessment and death or PD

The priority of the censoring rules is as follows:

1. If the subject had PD or death, the following sequence will be applied:
    - If a subject did not have a baseline tumor assessment (No. 1), the subject will be censored on the date of randomization. However, if the subject died within 125 days (18 weeks – 1 day) after randomization and did not receive new anticancer treatment, the date of death will be the PFS event date (not censored).
    - If a subject had new anticancer treatment before PD or death (No. 4), the subject will be censored on the date of the last tumor assessment prior to or on the start date of new anticancer treatment.
    - If PD is reported after the drug discontinuation visit, the subject will be censored at the date of last radiologic assessment before drug discontinuation except when PD occurred at the same visit as drug discontinuation (No. 8). If PD occurred within 4 weeks of drug discontinuation, it will be considered PD at the drug discontinuation visit and counted as a PFS event.
    - If a subject missed more than 1 assessment before PD or death (No. 7), the subject will be censored on the date of the last tumor assessment before PD or death. Note that if a subject is censored by both this criterion and the anticancer treatment criteria, the earliest censoring date will be used.
    - Otherwise, if a subject had an event (No. 2, No. 5 or No. 6), the earliest event date will be used.
2. If a subject did not have PD or death, the censoring date will be the earliest censoring date if the subject met multiple censoring criteria (No. 1, No. 3, No. 4, or No. 7).

8.5 Missing Date Handling

Missing data will not be imputed unless otherwise stated in the SAP or in the footnote of an output. Missing data will not be included in any analysis unless otherwise specified. When relevant, the number of subjects with missing data will be presented.

8.5.1 Adverse Event with Missing Dates

If the missing start and/or end dates of the AE do not indicate that the AE started prior to the start of study drug or after the 28th day after the last dose of study drug, it will be classified conservatively as a TEAE (Section 5.6.2). This data handling is only for the determination of TEAEs; no imputation will be performed to AE start/end dates in datasets and listings.

8.5.2 Medication with Date Missing or Partially Missing

See Section 5.2.5 for the missing date handling in determination of prior and concomitant medications.

8.6 Unscheduled Visits

Data from unscheduled visits (e.g., vital signs or laboratory tests) will be excluded from the by-visit summary but will contribute to the worst value in required summary tables. Listings will include all visits, both scheduled and unscheduled.

9 PROGRAMMING SPECIFICATIONS

The rules for programming derivations and dataset specifications will be provided in a separate document (analysis datasets specification).

10 STATISTICAL SOFTWARE

All statistical analyses will be performed using SAS version 9.3 or higher.

11 MOCK TABLES, LISTINGS, AND GRAPHS

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

12 REFERENCES

Common Terminology Criteria for Adverse Events (CTCAE). Version 4.03. United States Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, USA, June 14, 2010.

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

13 APPENDICES

13.1 Clinically Significant TEAEs

The following are the Clinically Significant AE groups considered in analyses:

1. Arterial Thromboembolic Events
2. Cardiac Dysfunction
3. Fistula Formation
4. GI Perforation
5. Hemorrhage
6. Hepatotoxicity
7. Hypertension
8. Hypocalcemia
9. Hypothyroidism
10. Palmar Plantar Erythrodysesthesia Syndrome
11. Posterior Reversible Encephalopathy Syndrome
12. Proteinuria
13. QT Prolongation
14. Renal Events

13.2 Thyroid Cancer Tumor-Node-Metastasis Staging System

The TNM (tumor-node-metastasis) Staging System is the most widely used system for cancer staging in the world. Created by the American Joint Committee on Cancer (AJCC), a distinguished group of experts from national healthcare organizations and major cancer centers around the country, the system defines cancer stage by describing:

Primary Tumor[a]	
Tx	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Tumor 2 cm or less in greatest dimension limited to the thyroid
T2	Tumor more than 2 cm but not more than 4 cm in greatest dimension limited to the thyroid
T3	Tumor more than 4 cm limited to the thyroid or any tumor with minimal extrathyroid extension (eg, extension to sternothyroid muscle or perithyroid soft tissues)
T4a	Tumor of any size extending beyond the thyroid capsule to invade subcutaneous soft tissues, larynx, trachea, esophagus, or recurrent laryngeal nerve
T4b	Tumor invades prevertebral fascia or encases carotid artery or mediastinal vessels
Regional Nodes[b]	
Nx	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis
N1a	Metastasis to Level VI (pretracheal, paratracheal, and prelaryngeal/Delphian lymph nodes)
N1b	Metastasis to unilateral, bilateral, or contralateral cervical or superior mediastinal lymph nodes
Distant Metastasis	
Mx	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis

Papillary or Follicular Carcinoma			
Stage Group[c]	**T Stage**	**N Stage**	**M Stage**
Under 45 Years			
I	Any T	Any N	M0
II	Any T	Any N	M1
45 Years and Older			
I	T1	N0	M0
II	T2	N0	M0
III	T3	N0	M0
	T1	N1a	M0
	T2	N1a	M0
	T3	N1a	M0
IVA	T4a	N0	M0
	T4a	N1a	M0
	T1	N1b	M0
	T1	N1b	M0
	T2	N1b	M0
	T3	N1b	M0
	T4a	N1b	M0
IVB	T4b	Any N	M0
IVC	Any T	Any N	M1

a. All categories may be subdivided: (a) solitary tumor, (b) multifocal tumor (the largest determines the classification).

b. Regional nodes are the central compartment, lateral cervical, and upper mediastinal lymph nodes.

c. Separate stage groupings are recommended for papillary or follicular, medullary, and anaplastic (undifferentiated) carcinoma.

13.3 Response Evaluation Criteria in Solid Tumors 1.1

Tumor response assessments in this clinical trial will use Response Evaluation Criteria in Solid Tumors (RECIST 1.1) based on the 2009 article by Eisenhauer, et al. entitled, *New Response Evaluation Criteria in Solid Tumors: revised RECIST guideline* (version 1.1).

The sole modification to RECIST 1.1 to be implemented in this study is that chest x-rays may not be used to follow disease; only CT scans may be used to follow chest disease. As required by RECIST 1.1, the protocol states that the minimum duration of SD is 7 weeks following the date of first dose of study drug.

The Eisenhauer article, published in the European Journal of Cancer, is available online at: http://linkinghub.elsevier.com/retrieve/pii/S0959804908008733.

13.4 Eastern Cooperative Oncology Group Performance Status

Scale	ECOG Performance Status
0	Fully active, able to carry on all predisease performance without restriction.
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature (eg, light house work, office work).
2	Ambulatory and capable of all self-care, but unable to carry out any work activities. Up and about more than 50% of waking hours.
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours.
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair.
5	Dead.

ECOG = Eastern Cooperative Oncology Group.
Adapted from Oken MM, et al. Am J Clin Oncol. 1982;5:649-55.

13.5 New York Heart Association Cardiac Disease Classification

The NYHA cardiac disease classification provides a functional and therapeutic classification for the prescription of physical activity for patients with heart failure, based on cardiac functional capacity. Based on NYHA definitions, subjects with heart failure are to be classified as follows:

Class	NYHA Status
Class I	Subjects with no limitation of activities; they suffer no symptoms from ordinary activities.
Class II	Subjects with slight, mild limitation of activity; they are comfortable at rest or with mild exertion.
Class III	Subjects with marked limitation of activity; they are comfortable only at rest.
Class IV	Subjects who should be at complete rest, confined to bed or chair; any physical activity brings on discomfort and symptoms occur at rest.

NYHA = New York Heart Association.

Adapted from The Criteria Committee of the New York Heart Association. *Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels*. 9th ed. 1994:253-6.

13.6 Common Terminology Criteria for Adverse Events (v4.03)

The NCI's Common Terminology Criteria for Adverse Events (CTCAE) v4.0 [published 28 May 2009 (v4.03: 14 June 2010)] provides descriptive terminology to be used for adverse event reporting in clinical trials. A brief definition is provided to clarify the meaning of each AE term. To increase the accuracy of AE reporting, all AE terms in CTCAE version 4.0 have been correlated with single-concept, Medical Dictionary for Regulatory Activities (MedDRA®) terms.

CTCAE v4.0 grading refers to the severity of the AE. CTCAE grades 1 through 5, with unique clinical descriptions of severity for each AE, are based on the general guideline for activities of daily living:

Grade	CTCAE Status
1	Mild: asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
2	Moderate: minimal, local, or noninvasive intervention indicated; limiting age-appropriate instrumental ADL.[a]
3	Severe or medically significant but not immediately life-threatening: hospitalization or prolongation of hospitalization indicated; disabling, limiting self-care ADL.[b]
4	Life-threatening consequences: urgent intervention indicated.
5	Death related to adverse event.

ADL = activities of daily living, CTCAE = Common Terminology Criteria for Adverse Events.
a: Instrumental ADL refers to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.
b: Self-care ADL refers to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden.
Adapted from the Cancer Therapy Evaluation Program, NCI. CTCAE v4.0. Available from: http://evs.nci.nih.gov/ftp1/CTCAE/About.html (Accessed 25 Jun 2015).

For further details regarding MedDRA, refer to the MedDRA website at: http://www.meddra.org/.

CTCAE v4.03 is available online at: http://evs.nci.nih.gov/ftp1/CTCAE/About.html (Accessed 25 Jun 2015).

SIGNATURE PAGE