STATISTICAL ANALYSIS PLAN

Study Protocol Number:	**E7080-G000-230**
Study Protocol Title:	A Multicenter, Open-label, Randomized Phase 2 Study to Compare the Efficacy and Safety of Lenvatinib in Combination with Ifosfamide and Etoposide versus Ifosfamide and Etoposide in Children, Adolescents and Young Adults with Relapsed or Refractory Osteosarcoma (OLIE)
Date:	15 June 2022
Version:	Final Version **3.0**

1 TABLE OF CONTENTS

2 VERSION HISTORY

Version	Description	Date
1.0	Initial Version	15Apr2020
2.0	Version based on Protocol Amendment 02 - Revised primary objective and efficacy endpoint to progression-free survival (PFS), and changed PFS rate at 4 months (PFS-4m) from primary to secondary objective and endpoint - The primary efficacy endpoint of PFS per IIR will be analyzed by the stratified log-rank test adjusting for the randomization stratification factors of "Age (<18 vs ≥18 years)" and "Time to first relapse/refractory disease (early [<18 months] vs late [≥18 months])". - Clarify that tumor efficacy endpoint(s) will be based on investigator and IIR evaluations. - Revised sample size calculation based on PFS, with the primary analysis being performed after 38 PFS events occur. - Add objective response rate (ORR) as a secondary endpoint. - Add 2 sensitivity analyses: 1) baseline tumor removal is considered part of the drug therapy, 2) PD or death after two or more missing assessments, or after receiving a new anticancer therapy will be considered as a PFS event. - Clarify exploratory endpoints - Revise the definition of the Randomization Phase to: The Randomization Phase will begin at the time of randomization of the first subject and will end on the data cutoff date for the primary analysis (ie, when approximately 38 PFS events, as determined by IIR, are observed). - Revise the length of follow-up from "will continue as long as the subject is alive" to "will continue for up to 2 years after end of treatment for a subject" unless the subject withdraws consent, or is lost to follow-up, or the sponsor terminates the study. - Added Optional Lenvatinib Crossover Treatment for Subjects in Arm B who experience progressive disease.	13Dec2021
3.0	- Updates to handling of subjects with removal of baseline lesions within 1) primary efficacy endpoint, 2) sensitivity analysis 1. - The primary efficacy endpoint of PFS per IIR will be analyzed by the stratified log-rank test with randomization	06Jun2022

	stratification factor of Age (<18 vs ≥18 years). Stratification factor of time to first relapse/refractory disease (early [<18 months] vs late [≥18 months]) is removed due to small subject numbers for stratum ≥ 18 months based on blinded data review.
	- Addition of subgroup analyses for the primary efficacy endpoint, PFS:
	  - Age (<18 years, ≥18 years)
	  - Time to first relapse (<18 months, ≥18 months)
	  - Sites of metastasis [Lung only, Others (+/- lung)]
	  - Prior Ifosfamide +/- Etoposide (Yes, No)
	  - Measurable disease [Yes (Subjects with at least 1 target lesion), No (Subjects with only non-target lesions)]
	  - Multiple editorial updates
	- For DCR, change SD minimum duration period to 5 weeks from 7 weeks after randomization. The change was made due to the tumor assessment schedule of every 6 weeks until Week 18.	

3 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AE	adverse event
ATC	anatomical therapeutic class
BLQ	below limit of quantification
BMI	body mass index
BSA	body surface area
CBR	clinical benefit rate
CMQ	customized MedDRA queries
CR	complete response
CRF	case report form
CSAEs	clinically significant adverse events
CTCAE	Common Terminology Criteria for Adverse Events
DCR	disease control rate
DOR	duration of response
DBP	diastolic blood pressure
EMA	European Medicines Agency
HRQoL	Health-Related Quality of Life
IIR	independent imaging review
IRT	interactive response technology
ITT	intent-to-treat
K-M	Kaplan-Meier
KPS	Karnofsky performance status
LLT	lower level term
LVEF	left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
NA	not applicable
NE	not evaluable
ORR	objective response rate
OS	overall survival
OS-1y	overall survival rate at 1 year

Abbreviation	Term
PD	pharmacodynamic(s) or progressive disease/disease progression
PedsQL	Pediatric Quality of Life Inventory
PFS	progression-free survival
PFS-4m	progression-free survival rate at 4 months
PFS-1y	progression-free survival rate at 1 year
PK	pharmacokinetic(s)
PR	partial response
PRO	patient-reported outcome
PT	preferred term
QTc	corrected QT interval
QTcB	corrected QT interval using Bazett's correction factor
QTcF	corrected QT interval using Fridericia's correction factor
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation or stable disease
SI	Système International
SMQs	standardized MedDRA queries
SOC	system organ class
TEAE	treatment-emergent adverse event
TLG	tables, listings, and graphs
TNM	tumor-node-metastasis
TSH	thyroid stimulating hormone
WHO DD	World Health Organization Drug Dictionary

4 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-G000-230 Amendment 02. Analyses on subjects' Health-Related Quality of Life (HRQoL) data, pharmacokinetic (PK)/pharmacodynamics (PD), biomarkers, and relationships between PK and efficacy/safety will be included in separate statistical analysis plans.

4.1 Study Objectives

4.1.1 Primary Objective

To evaluate whether lenvatinib in combination with ifosfamide and etoposide (Arm A) is superior to ifosfamide and etoposide (Arm B) in improving progression-free survival (PFS) by independent imaging review (IIR) using Response Evaluation Criteria In Solid Tumors (RECIST 1.1 [Eisenhauer, et al., 2009]), in children, adolescents, and young adults with relapsed or refractory osteosarcoma.

4.1.2 Secondary Objectives

The secondary objectives of the study are to:

1. Compare difference in PFS rate at 4 months (PFS-4m) between the 2 treatment arms per IIR
2. Compare difference in PFS rate at 1 year (PFS-1y) between the 2 treatment arms per IIR
3. Compare differences in overall survival (OS) and OS rate at 1 year (OS-1y) between the 2 treatment arms
4. Compare difference in objective response rate (ORR) at 4 months (ORR-4m) between the 2 treatment arms per IIR
5. Compare difference in objective response rate (ORR) between the 2 treatment arms per IIR
6. Compare differences in safety and tolerability between the 2 treatment arms
7. Characterize the PK of lenvatinib, when administered in combination with ifosfamide and etoposide
8. Compare differences in HRQoL as assessed by using the Pediatric Quality of Life Inventory (PedsQL) Generic Core Scales and Cancer Module between the 2 treatment arms
9. Assess the palatability and acceptability of the suspension formulation of lenvatinib in subjects receiving the suspension formulation in the study

4.1.3 Exploratory Objectives

The exploratory objectives of the study are to:

1. Explore differences in duration of response (DOR), disease control rate (DCR), and clinical benefit rate (CBR) between the 2 treatment arms per IIR and investigator assessment+
2. Explore differences in PFS, PFS-4m, PFS-1y, ORR-4m, and ORR between the 2 treatment arms per investigator assessment
3. Compare between the 2 treatment arms:
   - the proportion of subjects who achieve complete removal of baseline lesion(s)
   - the proportion of subjects with unresectable baseline lesion(s) that are converted to resectable
4. Investigate the relationship between subject tumor biomarkers and clinical response and toxicity of lenvatinib in combination with ifosfamide and etoposide

4.2 Overall Study Design and Plan

This is a multicenter, randomized, open-label, parallel-group, Phase 2 study to compare the efficacy and safety of lenvatinib in combination with ifosfamide and etoposide versus ifosfamide and etoposide in children, adolescents, and young adults with relapsed or refractory osteosarcoma.

Approximately 72 eligible subjects will be randomized to 1 of the following 2 treatment arms in a 1:1 ratio within the strata:

- Arm A: lenvatinib 14 mg/m^2 (orally, once daily) plus ifosfamide 3000 mg/m^2/day (intravenously [IV], Day 1 to Day 3 of each cycle for a total of up to 5 cycles) and etoposide 100 mg/m^2/day (IV, Day 1 to Day 3 of each cycle for a total of up to 5 cycles)
- Arm B: ifosfamide 3000 mg/m^2/day (IV, Day 1 to Day 3 of each cycle for a total of up to 5 cycles) and etoposide 100 mg/m^2/day (IV, Day 1 to Day 3 of each cycle for a total of up to 5 cycles)

Randomization will follow a predefined randomization scheme based on the following stratification factors: time to first relapse/refractory disease (early [<18 months] or late [≥18 months]) and age (<18 and ≥18 years).

Eisai will closely monitor enrolment, to ensure that at least 36 subjects are <17 years of age at the time of informed consent per regulatory requirement.

The study will be conducted in 3 Phases: a Prerandomization Phase, a Randomization Phase, and an Extension Phase.

The end of the study will be the date of data cutoff for the final analysis or the time of last subject last visit, whichever occurs later.

An overview of the study design is presented in Figure 1 in the Protocol.

4.2.1 Prerandomization Phase

The Prerandomization Phase will consist of 2 periods: Screening and Baseline. The Prerandomization Phase will last no longer than 28 days. The Screening Period will establish protocol eligibility and the Baseline Period will confirm eligibility.

4.2.1.1 Screening Period

Screening will occur between Day –28 and Day –2. The purpose of the Screening Period is to obtain informed consent and to establish protocol eligibility. Informed consent will be obtained after the study has been fully explained to each subject and before the conduct of any screening procedures or assessments. Procedures to be followed when obtaining informed consent are detailed in Section 5.3 of the Protocol.

Subjects must have a histologically or cytologically confirmed diagnosis of high grade refractory or relapsed osteosarcoma as detailed in the Inclusion Criteria (Section 9.3.1 of the Protocol).

The Screening Disposition case report form (CRF) page must be completed to indicate whether the subject is eligible to participate in the study and to provide reasons for screen failure, if applicable.

4.2.1.2 Baseline Period

The purpose of the Baseline Period is to confirm protocol eligibility as specified in the inclusion/exclusion criteria (as detailed in Section 9.3.1 and Section 9.3.2 of the Protocol). Results of baseline assessments must be obtained prior to the first dose of study drug (Cycle 1 Day 1). Baseline assessments may be performed on Day -1 or on Cycle 1 Day 1 prior to dosing. Clinical laboratory tests (Table 4 of the Protocol), including a pregnancy test (where applicable), should be performed within 72 hours prior to the first dose of study drug.

Subjects who complete the Baseline Period and continue to meet the criteria for inclusion/exclusion (as detailed in Section 9.3.1 and Section 9.3.2 of the Protocol) will begin the Randomization Phase of this study.

4.2.2 Randomization Phase

The Randomization Phase will consist of 2 periods: Treatment Period and Follow-up Period. The Randomization Phase will begin at the time of randomization of the first subject and will end on the data cutoff date for the primary analysis (ie, when approximately 38 PFS events, as determined by IIR, are observed).

After the data cutoff date for the primary analysis has occurred, all subjects who are still on study treatment will enter the Extension Phase.

4.2.2.1 Treatment Period

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

Subjects will receive study treatment as continuous 21-day cycles in both treatment arms. Treatment cycles will be counted continuously regardless of dose interruptions. If chemotherapy administration is precluded on Day 1 of a treatment cycle, the subject may resume treatment once recovered at the discretion of the investigator. Subsequent chemotherapy will be administered every 21 (±1) days starting from the timepoint it was resumed. Subjects will undergo safety and efficacy assessments as defined in the Schedule of Procedures/Assessments (Table 5 of the Protocol). Subjects randomized to Arm A will continue to receive lenvatinib until disease progression (PD) confirmed by IIR, development of unacceptable toxicity, subject request, withdrawal of consent, or study termination by the sponsor, whichever occurs first.

Disease progression (PD) must be confirmed by IIR prior to the investigator discontinuing study treatment for a subject. In situations where the investigator judges that alternative treatments must be instituted immediately for a subject's safety, study drugs may be discontinued without waiting for IIR confirmation of radiographic evidence of PD. If possible, before discontinuation of the subject from the study, the investigator should consult with the sponsor.

4.2.2.2 Follow-up Period

The Follow-up Period begins the day after the Off-Treatment Visit and will continue for up to 2 years after end of treatment for a subject, unless the subject withdraws consent, or is lost to follow-up, or the sponsor terminates the study.

Subjects may at any time withdraw consent for further study participation. No further data will be collected on subjects once consent has been withdrawn; however, an investigator may consult public records to establish survival status if permitted by local regulations.

All adverse events (AEs), including serious adverse events (SAEs), will be captured for 30 days after the last dose of study drug.

All subjects who end study treatment without IIR confirmed PD will continue to undergo tumor assessments every 6 weeks until Week 18, then every 9 weeks until Week 54, thereafter, every 12 weeks until IIR confirmation of radiographic evidence of PD as described in the tumor assessments in the assessment schedule, or until another anticancer therapy is initiated.

Subjects in both Arm A and Arm B will be followed for survival every 12 weeks (±1 week) and all subsequent anticancer treatments received will be recorded. Subjects who are being

followed for survival at the primary analysis (ie, at the end of the Randomization Phase) will continue to be followed for survival during the Follow-up Period of the Extension Phase.

4.2.3 Extension Phase

The Extension Phase starts after completion of the Randomization Phase (data cutoff for primary analysis). The Extension Phase will consist of 2 periods: Treatment Period and Follow-up Period.

4.2.3.1 Treatment Period

In the Treatment Period, subjects still on study treatment following the completion of the Randomization Phase will continue study treatment as outlined in Protocol Section 9.4.1 until disease progression, development of unacceptable toxicity, subject request, withdrawal of consent, or discontinuation of study by the sponsor. Tumor assessments will be performed according to the local standard of care. Independent imaging review (IIR) and confirmation of radiographic evidence of PD will not be required, and scans will no longer be required to be sent to the imaging core laboratory. The Off-Treatment Visit will occur within 30 days after the final dose of study treatment. All AEs, including SAEs will be captured up to 30 days after last dose of study drug. In case the study is discontinued by the sponsor, the sponsor will continue to provide study drug (outside the study) for subjects who have not met the criteria for study drug discontinuation.

4.2.3.2 Follow-Up Period

The Follow-up Period will begin the day after the Off-Treatment Visit and will last for up to 2 years after end of treatment for a subject, unless the study is terminated by the sponsor, or the subject discontinues due to withdrawal of consent, or is lost to follow-up. Subjects will be treated by the investigator according to the prevailing local standard of care. Subjects will be followed every 12 weeks (±1week) for survival and all subsequent anticancer treatments received will be recorded.

4.2.4 Optional Lenvatinib Crossover (for Subjects in Arm B Only)

Subjects in Arm B with disease progression per RECIST 1.1 may be eligible for optional treatment with lenvatinib alone (after 5 cycles of chemotherapy) or lenvatinib plus chemotherapy (before 5 cycles of chemotherapy). Note: subjects may only receive a maximum of 5 cycles of chemotherapy for the duration of the study. Optional Lenvatinib crossover treatment must be initiated within 30 days of documented disease progression.

Optional lenvatinib crossover treatment is only available if the following conditions are met:

- Subject experiences disease progression per RECIST 1.1 (as confirmed by IIR for all subjects who cross over prior to the start of the Extension Phase); and
- No new systemic anticancer medication was administered after the last dose of study drugs; and
- The subject meets all of the safety parameters listed in the inclusion criteria and none of the safety parameters listed in the exclusion criteria; and

- The study is ongoing.

Treatment with lenvatinib alone or lenvatinib plus chemotherapy will continue until the next disease progression (per RECIST 1.1 as assessed by investigator only), development of unacceptable toxicity, subject request, or withdrawal of consent, whichever occurs first.

Prior to optional lenvatinib crossover treatment, baseline tumor assessment must be re-established (ie, new tumor assessment scans performed), unless the last tumor assessment scans were performed within 4 weeks prior to Cycle 1 Day 1 of the crossover treatment.

Subjects who qualify to receive optional lenvatinib crossover treatment will be followed according to the schedule of procedures/assessments in Table 7 of the Protocol.

5 DETERMINATION OF SAMPLE SIZE

A total sample size of 72 subjects is estimated for the primary efficacy endpoint of PFS. A median PFS of 3.5 months for Arm B (control arm) has been assumed by comparing available PFS data from clinically active agents in this patient population (Davis, et al., 2019; Palmerini, et al, 2016; Grignani, et al., 2012). With respect to Arm A (lenvatinib plus chemotherapy arm), a median PFS of 8.75 months is assumed based on the results from Study E7080-G000-207.

Therefore, assuming a hazard ratio of 0.4, a 1-sided type 1 error rate of 0.025, and power of 80%, the required number of PFS events for the study is 38. The primary PFS analysis is estimated to occur at approximately 32 months after the first subject is randomized (assuming a 20-month enrollment period), and accounts for an overall dropout rate of up to 40%. This dropout rate includes the potential early dropouts and subjects who are censored before the data cutoff.

6 STATISTICAL METHODS

In general, continuous variables will be summarized using descriptive statistics such as mean, SD, median, first quartile (Q1), third quartile (Q3), and range (minimum and maximum). Categorical variables will be summarized using frequency and percentage. For time-to-event variables, the Kaplan-Meier (K-M) method will be used for descriptive summaries. For the calculation of time-to-event or duration-of-event variables, the difference between the start date and the end date plus 1 day will be used. For 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 9.4.1.

For any efficacy endpoint that involves complete response (CR) or partial response (PR), the endpoint will be calculated using the confirmed CR/PR, i.e., any CR or PR must be confirmed not less than 4 weeks following the initial achievement of the response.

6.1 Study Endpoints

Efficacy endpoints related to tumor assessments will be evaluated by IIR and investigator assessment during the randomization phase and investigator assessment only during the extension phase.

6.1.1 Primary Endpoint

Progression-free survival (PFS) by IIR is defined as the time from the date of randomization to the date of the first documentation of PD or death (whichever occurs first) as determined by IIR using RECIST 1.1.

6.1.2 Secondary Endpoints

1. Progression-free survival rate at 4 months (PFS-4m) by IIR is defined as the percentage of subjects who are alive and without PD at 4 months from the randomization date as determined by IIR of radiological imaging using RECIST 1.1. The PFS-4m rate is estimated using the Kaplan-Meier (K-M) method.
2. Progression-free survival rate at 1 year (PFS-1y rate) by IIR is defined as the percentage of subjects who are alive and without PD at 1 year from the randomization date as determined by IIR of radiological imaging using RECIST 1.1. The PFS-1y rate is estimated using the K-M method.
3. Overall survival (OS) is defined as the time from the date of randomization to the date of death from any cause. Subjects who are lost to follow-up and those who are alive at the date of data cutoff for the primary analysis, will be censored at the date the subject was last known to be alive, or date of data cutoff for the primary analysis, whichever occurs first. Overall survival rate at 1 year (OS-1y) will be estimated using the K-M method.
4. Objective response rate by IIR at 4 months (ORR-4m) is defined as the proportion of subjects who have best overall response of complete response (CR) or partial response (PR) as determined by IIR using RECIST 1.1 within the first 4 months.
5. Objective response rate (ORR) by IIR is defined as the proportion of subjects who have best overall response of CR or PR as determined by IIR using RECIST 1.1.
6. Safety will be assessed by summarizing the incidence of treatment-emergent adverse events (TEAEs) and SAEs together with all other safety parameters.
7. Assessment of population-based PK parameters of lenvatinib.
8. Score changes from baseline for all PedsQL scales including Generic Core Scales and Cancer Module. Scores will be calculated for total generic score, total cancer score, each physical function subscale including physical health, psychosocial health, emotional function, social function, school/work function in the Generic Core Scales, and each subscale in the cancer module.
9. Palatability and acceptability of the suspension formulation of lenvatinib in subjects receiving the suspension formulation in the study will be assessed by using the Palatability Questionnaire (see Appendix 4 of the Protocol).

6.1.3 Exploratory Endpoints

1. Duration of response (DOR) by IIR and investigator assessment is defined as the time from the date a response was first documented until the date of the first documentation of PD or date of death from any cause.
2. Disease control rate (DCR) by IIR and investigator is the proportion of subjects who have a best overall response of CR or PR or stable disease (SD). In this context, stable disease is defined as stable disease at ≥5 weeks after randomization to be considered best overall response.
3. Clinical benefit rate (CBR) by IIR and investigator is the proportion of subjects who have best overall response of CR or PR or durable SD (duration of SD ≥23 weeks after randomization).
4. Efficacy endpoints (PFS, PFS-4m, PFS-1y, ORR-4m, and ORR) evaluated based on investigator assessment.
5. Proportion of subjects who achieve complete removal of baseline lesions and the proportion of subjects with unresectable baseline lesions(s) that are converted to resectable between the 2 treatment arms.
6. Blood and tumor biomarkers will be assessed for identifying potential correlation with clinical outcomes-related endpoints.

6.2 Study Subjects

6.2.1 Definitions of Analysis Sets

- The Full Analysis Set (Intent-to-Treat Analysis [ITT]) includes all randomized subjects regardless of the treatment actually received. This is the primary analysis population used for the efficacy analyses which will be based on the ITT principle.
- The Per Protocol Analysis Set includes those subjects from the ITT set who received at least 1 dose of any study drug, had no important protocol deviations, and had both baseline and at least one postbaseline tumor assessment. Subjects for whom death occurred prior to the first postbaseline tumor assessment will also be included. The per protocol analysis set will be the secondary analysis set for the primary efficacy endpoint.
- The Safety Analysis Set includes subjects who received at least 1 dose of any study drug. This is the analysis population used for all safety analyses which will be based on the as-treated principle.
- Population Pharmacokinetic (PK) Analysis Set includes the subjects who have received at least 1 dose of lenvatinib with documented dosing history and have measurable plasma levels of lenvatinib.
- The Pharmacodynamic Analysis Set includes subjects who received at least 1 dose of study drug and had sufficient pharmacodynamic data (eg, at least 1 evaluable/measurable pharmacodynamic parameter).
- The HRQoL Analysis Set will consist of all randomized subjects who have received at least 1 dose of study medication, and have completed at least 1 patient-reported outcome (PRO) assessment beyond baseline. For the analysis, subjects will be analyzed as randomized and not according to treatment actually received.

Analysis sets in data analyses are tabulated below.

Table 1 Analysis Sets in Data Analysis

Tables	All Screened Subjects	Full Analysis Set	Per-Protocol Analysis Set	Safety Analysis Set	PK Analysis Set	PD Analysis Set	HRQoL Analysis Set
Protocol Deviations		●					
Disposition	●	●					
Demography & Baseline Characteristics		●					
Disease History		●					
Prior and concomitant medication		●					
New anticancer therapy		●					
Efficacy analysis		●	●				
Safety analysis (including drug exposure, AEs, laboratory tests, vital signs, etc.)				●			
Pharmacokinetics					●		
Pharmacodynamics						●	
HRQoL							●

AEs = adverse events, HRQoL = Health-Related Quality of Life, PD = pharmacodynamic, PK = pharmacokinetic.

6.2.2 Subject Disposition

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 summary of subject disposition will be conducted on the Full Analysis Set.

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

The end of study status (Alive, Death, Withdrew Consent, Lost to Follow-up, etc.) at the data cutoff date will also be summarized by treatment arm and overall.

6.2.3 Protocol Deviations

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

Important protocol deviations will be appropriately grouped into different categories. They will be summarized by treatment arm within the Full Analysis Set if necessary. A list of subjects with important protocol deviations will be provided.

In addition, a listing of all protocol deviations due to COVID-19 will be provided.

6.2.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics will be summarized for the Full Analysis Set.

The following demographic and baseline characteristics will be summarized:

- Age (years)
- Age group (≥2 years to <6 years, ≥6 years to <17 years [≥6 years to <12 years, ≥12 years to <17 years], ≥17 years)
- Age group (≥2 years to <12 years, ≥12 years to <18 years, ≥18 years)
- Age group from interactive response technology (IRT) (<18 and ≥18 years)
- Gender (Male, Female)
- Race (White, Black, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other); Asian subcategories: Chinese, Japanese, Other.
- Race group (White, Non-White)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Region (North America, Europe, and Asia Pacific)
- Pregnancy test results (Positive, Negative) in females
- Tanner stage
- Height (cm)
- Baseline weight (kg)
- Baseline body mass index (BMI) (kg/m^2)
- Baseline body surface area (BSA) (m^2)
- Karnofsky performance status (KPS) or Lansky play scores (LPS)
- Strata will be summarized based on IRT and CRF data. Strata are defined by the two stratification factors of time to first relapse/refractory disease and age:
  - Early (<18 months) and age group (<18 years)
  - Early (<18 months) and age group (≥18 years)
  - Late (≥18 months) and age group (<18 years)
  - Late (≥18 months) and age group (≥18 years)

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

- Time since first diagnosis of osteosarcoma to randomization date (months)
- Time since metastatic diagnosis to randomization date (months)
- Time since last progression to randomization date (months)
- Age at diagnosis (in years)
- Osteosarcoma subtypes
- Location of the primary tumor
- TNM staging and grades
- Measurable disease [Yes (Subjects with at least 1 target lesion), No (Subjects with only non-target lesions)]
- Resectable Lesions at Baseline
- Sites of metastasis at baseline (bone, lung, bone AND lung, brain, liver, lymph, and other)

The following previous anticancer therapies/medications will also be summarized:

- Number of previous medication regimens
- Duration of most recent medication (months)
- Best response for most recent medication [complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), not evaluable (NE), not applicable (NA), unknown]
- Time from end of most recent medication to date of randomization (months)
- Therapeutic setting (Adjuvant, Maintenance, Neo-adjuvant, and Therapeutic)
- Previous radiotherapy (Yes, No)
    - Time from end of most recent radiotherapy to date of randomization (months)
    - Site of previous radiotherapy
    - Tumor lesion at the site progressed since most recent radiotherapy (Yes, No, Not evaluable)
- Previous anthracycline therapy (Yes, No)
- Previous receptor tyrosine kinase inhibitor (RTKI) therapy (Yes, No)
- Previous ifosfamide monotherapy (Yes, No)
- Previous ifosfamide plus etoposide (Yes, No)
- Previous Anticancer Procedures (Procedure Name; Site/Organ; Has tumor lesion at the site progressed since the procedure? (Yes, No); Post Op Result: Complete Resection, Incomplete Resection).

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

6.2.4.1 Medical History

General medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 25.0 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 PT, current medical condition, date of diagnosis or surgical procedure or onset of symptoms, and end date/ongoing.

6.2.5 Prior and Concomitant Therapy

All investigator terms for medications recorded on the CRF will be coded using the World Health Organization Drug Dictionary (WHO DD) Version of March 2022.

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 30 days after the subject's last dose. Medications that cannot be determined to be prior/concomitant 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 antidiarrheal therapy and antihypertensive therapy. Concomitant antidiarrheal therapy and antihypertensive therapy will be summarized separately.

Prior anticancer therapies including anticancer medications, anticancer procedures and anthracycline therapy will be summarized and listed.

Prior and concomitant medications/therapies will be summarized for the Full Analysis Set.

6.2.6 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 6.6.1.2; no other analysis of treatment compliance is planned.

6.3 Data Analysis General Considerations

6.3.1 Pooling of Centers

Subject data from all centers will be pooled for all analyses. Center will not be considered as a factor in the analyses.

6.3.2 Adjustments for Covariates

The stratified log-rank test for the primary efficacy endpoint, PFS as assessed by IIR, the secondary efficacy endpoint, OS, the stratified Miettinen and Nurminen's method for the secondary efficacy endpoint of ORR, and the stratified Cox regression analyses to estimate

hazard ratios (time-to-event efficacy endpoints) will be performed using the following randomization stratification factors from IRT.

- Time to first relapse/refractory disease (early [<18 months] or late [≥18 months])
- Age (<18 and ≥18 years)

In case of low numbers of subjects enrolled within any combination of stratification factors, subjects from the low enrollment combination(s) of the stratification factors may be pooled for analysis.

Based on a blinded review of the data, the number of subjects in the level of the stratum "time to first relapse/refractory disease ≥18 months" is approximately 11 subjects out of 81 subjects in the ITT population. To ensure a sufficient number of participants in each stratum within each treatment arm, "time to first relapse/refractory disease" has been removed as a stratification factor in the analyses. Thus, only the stratification factor "age at randomization" will be included in any stratified analyses.

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

6.3.3 Multiple Comparisons/Multiplicity

No multiplicity adjustment will be made for this study.

6.3.4 Examination of Subgroups

For the primary efficacy endpoint PFS as assessed by IIR, the median PFS, hazard ratio, and corresponding 95% confidence intervals may be generated for the following subgroups, as appropriate.

- Age (<18 years, ≥18 years)
- Time to first relapse (<18 months, ≥18 months)
- Sites of metastasis [Lung only, Others (+/- lung)]
- Prior Ifosfamide +/- Etoposide (Yes, No)
- Measurable disease [Yes (Subjects with at least 1 target lesion), No (Subjects with only non-target lesions)]

For the safety endpoints, the following TEAE tables will be provided by age group (years) category (<18 years, ≥18 years) and gender.

- Overall summary of TEAE
- TEAEs in decreasing frequency of PT
- Serious TEAEs in decreasing frequency of PT

Other exploratory subgroup analyses may be conducted if appropriate.

6.3.5 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.

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

Adverse events with incomplete dates will be considered treatment emergent if:

- Day and month are missing and the year is equal to or after the year of the first study drug dose date
- Day is missing, and the year is after the year of the first study drug dose
- Day is missing and the year is equal to the year of the first study drug dose date and the month is equal to or after the month of the first dose date
- Year is missing; or complete date is missing

HRQoL analysis will follow the FDA and the European Medicines Agency (EMA) PRO guidelines. Handling of missing values for HRQoL analyses will be detailed in a separate SAP and HRQoL report.

6.4 Efficacy Analyses

Efficacy analyses will be based primarily on the Full Analysis Set. The per protocol analysis set will be the secondary analysis set for efficacy endpoints. All tumor assessment analyses using the IIR assessments will be considered primary; all tumor assessment analyses using investigator assessments will be considered supportive analyses. The randomization strata based on IRT will be used in the analyses unless otherwise specified.

All primary statistical analyses will be conducted at the data cutoff date for the analysis of the primary efficacy endpoint, PFS by IIR (ie, when approximately 38 PFS events, as determined by IIR, are observed). Additional analyses will be based on the date of data cutoff for the additional follow-up analysis of OS or at the time of the last subject last visit, whichever occurs later.

6.4.1 Primary Efficacy Analysis

The primary efficacy endpoint of PFS per IIR will be analyzed and compared between the 2 treatment arms by the stratified log-rank test adjusted for the randomization stratification factor (refer to Section 6.3.2). Median PFS will be calculated using the K-M product-limit estimates for each treatment arm along with 2-sided 95% CIs (estimated with a generalized Brookmeyer and Crowley method [Brookmeyer and Crowley, 1982]). The K-M estimates of PFS will be plotted over time. The stratified Cox regression model will be used to estimate the hazard ratio and its 95% CI.

The PFS by IIR is defined as the time from the date of randomization to the date of first documentation of disease progression based on IIR assessments using RECIST 1.1 or date of death, whichever occurs first.

Determination of the date of PFS event or censoring is summarized in Table 2 below. The table is based on the FDA clinical trial endpoints guidance of 2007 (FDA, 2007). In this study, removal of baseline lesion(s) is permitted after completion of the Week 18 tumor assessment, and palliative radiotherapy of non-target lesions is allowed (see protocol 9.4.7.3). For the primary efficacy endpoint analysis, PFS will be censored according to censoring rules in Table 2. In this analysis, assessments after removal of baseline lesion(s) (considered as a new anticancer therapy) before or after Week 18 will be censored.

Table 2 PFS Event and Primary Censoring Rules

No.	Situation	Date of Progression or Censoring	Outcome
1	No baseline or post-baseline adequate tumor assessments*	Date of randomization	Censored
2	Progression documented between scheduled visits, on or prior to new anticancer therapy	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 without disease progression or any new anticancer therapy	Date of death	PFS Event
6	Death between adequate assessment* (Death before next scheduled assessment or after exactly one missing assessment)	Date of death	PFS Event
7	Death or progression after more than one consecutive missed visit/tumor assessment**	Date of last adequate radiologic assessment before missed tumor assessments	Censored

Note:

+ Assessments after removal of baseline lesion(s) (eg, surgical, radiofrequency ablation, cryotherapy, thermoablation, stereotactic radiotherapy, etc.) before or after Week 18 will be censored. Palliative radiotherapy of nontarget lesions is permitted per protocol and will not be considered as a new anticancer therapy.

* Adequate tumor assessment is radiologic assessment at regular intervals as defined in the protocol. Any tumor assessments occurring after the start of a new anticancer treatment will be censored per the FDA censoring rules.

** More than one consecutive missed visit/adequate tumor assessment is defined as the duration between two consecutive tumor assessments or the duration between the last adequate tumor assessment and death/PD is longer than (>) 97 days [= ((6+1) x 2 x 7) – 1] for subjects on every 6 weeks tumor assessment schedule in the first 18 weeks (6 cycles) of treatment; longer than (>) 139 days [= ((9+1) x 2 x 7) – 1)] for subjects on every 9 weeks tumor assessment schedule from weeks 18 to 54 (Cycle 7 to Cycle 18); longer than (>) 195 days [=((12+2) x 2 x 7) – 1)] for subjects on every 12 weeks tumor assessment schedule after Week 54 in this study.

For subjects moving from one schedule to another:

Missed Week 18 and Week 27 assessments: (>)118 days [= ([(6+1) + (9+1)] x 7) – 1)]

Missed Week 54 and Week 66 assessments: (>)167 days [= ([(9+1) + (12+2)] x 7) – 1)]

If a subject had PFS event (#2, #5 or #6), the earliest event date will be used.

The priority of the censoring rules is described as follows:

- If a subject missed two or more consecutive tumor assessments before PD or death (#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 this criterion and start of new anticancer therapy criterion, the earliest censoring date will be used.

- 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 (#1, #3, #4, #7).

Sensitivity Analysis of PFS

Sensitivity analyses of PFS will be based on the Full Analysis Set using censoring rules from Table 2, along with the following exceptions listed below:

Sensitivity Analysis 1: In this analysis, baseline tumor removal before or after completion of the Week 18 tumor assessment will be evaluated as part of the study drug therapy and will not be considered as a new anticancer therapy (see difference in Table 2, rule 4).

Sensitivity Analysis 2 (EMA guidance [Reference 11]): In this analysis, PD or death after two or more missing assessments, or after receiving a new anticancer therapy will be considered as a PFS event (see differences in Table 2, rules 4 and 7). The removal of baseline lesion(s) before or after completion of the Week 18 tumor assessment will be considered as a new anticancer therapy.

6.4.2 Secondary Efficacy Analyses

6.4.2.1 PFS-4m and PFS-1y Rate by IIR

The PFS-4m rate and PFS-1y rate per IIR will be estimated using the K-M method for the primary efficacy endpoint PFS per IIR. PFS-4m/PFS-1y rate and their Greenwood standard errors will be evaluated using the K-M estimates from both treatment groups. The statistical significance of the difference in the 2 K-M PFS-4m/PFS-1y rates will be based on the 2-sided 95% CI. This 2-sided 95% CI and a p-value will be constructed using the difference of these 2 K-M PFS-4m/PFS-1y rates and the 2 corresponding Greenwood standard errors.

6.4.2.2 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 who withdrew consent will be censored at the date the subject was last known to be alive. Subjects who are still alive at the data cut-off date will be censored at the data cut-off date.

Determination of date of OS event or censoring is summarized in Table 3 below.

Table 3 OS Event and Censoring Method

Situation	Date of Event or Censoring	Outcome
Death during study	Date of death	OS Event
Subject still alive at data cut-off (including death after 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

OS = overall survival.

Overall survival (OS) will be compared between the treatment arm and control arm using the stratified log-rank test adjusted for the randomization stratification factor (refer to Section 6.3.2).

Median will be calculated using the K-M [Kaplan and Meier, 1958] product-limit estimates for each treatment arm along with 2-sided 95% CIs (estimated with a generalized Brookmeyer and Crowley method [Brookmeyer and Crowley,1982]). The K-M estimates of OS will be plotted over time.

The Cox regression model will be used to estimate the hazard ratio and the corresponding 95% CIs stratified by the randomization stratification factor (refer to Section 6.3.2). For all

analyses using Cox regression, Efron's method will be used for handling "ties" in the failure time.

K-M estimates will also be presented for 4, 6, 9 and 12 months (ie, OS-1y) along with 2-sided 95% CIs (estimated with the Greenwood formula using the log-log transformation [Greenwood (1926) and Kalbfleisch and Prentice (2002)]). The OS-1y will be analyzed using the same approach as for the PFS-4m.

6.4.2.3 Overall Response Rate (ORR) and ORR-4m by IIR

The ORR by IIR is the proportion of subjects with a best overall response of CR or PR determined by IIR using RECIST 1.1. The ORR by IIR will be summarized and compared between the two treatment arms. The raw difference in ORR will be reported. The 2-sided 95% confidence interval of ORR difference will be generated from SAS using Miettinen-Nurminen (Score) Confidence Limits.

The ORR-4m by IIR will be summarized and compared between arms using the same approach as for ORR by IIR.

6.4.2.4 Duration of Response, Disease Control Rate, and Clinical Benefit Rate by IIR

Median DOR by IIR among responders for each arm will be presented along with the corresponding 2-sided 95% CI; 95% CI will be estimated with the Brookmeyer and Crowley method.

Disease Control Rate (DCR) and CBR by IIR will be summarized and compared between arms using the same approach as for ORR.

6.4.2.5 PFS, PFS-4m, PFS-1y, ORR-4m, ORR, DOR, DCR, and CBR by Investigator Assessment

Efficacy endpoints PFS, PFS-4m, PFS-1y, ORR-4m, ORR, DOR, DCR, and CBR evaluated based on investigator assessment will be analyzed using the same approach as for the endpoints assessed by IIR.

6.4.2.6 Subjects Achieving Complete Resection

The difference in the proportion of subjects who achieve complete removal of baseline lesion(s) between the 2 treatment arms will be compared and corresponding two-sided 95% CIs will be calculated using the Clopper and Pearson method (Clopper and Pearson, 1934). The proportion of subjects with unresectable baseline lesions(s) that are converted to resectable lesions will be compared between the 2 treatment arms in the same way.

6.4.2.7 Maximum Tumor Shrinkage

Graphical displays (waterfall plot) will be prepared for the maximum tumor shrinkage in target lesions, defined as the maximum percentage change from baseline in the sum of the diameters of the target lesions at data cutoff.

6.5 Pharmacokinetic, Pharmacodynamic, Pharmacogenomic, and Other Biomarker Analyses

6.5.1 Pharmacokinetic Analyses

Lenvatinib plasma concentration data will be summarized across timepoints using n, mean, SD, median, Q1, Q3, minimum, and maximum. Graphical displays will present scatter plots and box plots of lenvatinib plasma concentration over time. Plasma concentration of lenvatinib versus time data will be listed.

Plasma concentration data of lenvatinib will be analyzed using a population PK approach to estimate population PK parameters. The analysis will be detailed in a separate analysis plan.

6.5.2 Pharmacodynamic, Pharmacogenomic, and Other Biomarker Analyses

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

6.6 Safety Analyses

All safety analyses will be performed on the Safety Analysis Set. Safety data, presented by treatment group, will be summarized on an "as treated" basis using descriptive statistics (eg, n, mean, standard deviation, median, minimum, maximum, Q1, and Q3 for continuous variables; n [%] for categorical variables), as appropriate. All safety data will be summarized up to the last dose + 30 days or the data cutoff date, whichever occurs earlier. Optional lenvatinib crossover treatment is permitted by protocol. For subjects in Arm B who crossed over to optional treatment with lenvatinib alone or lenvatinib plus chemotherapy, data records occurring on or after the date of the crossover (date of first administration of lenvatinib) will be excluded from this analysis. In addition, for these subjects, data occurring on or after the date of the crossover will be summarized and listed separately for selected parameters as appropriate.

Safety variables include TEAEs, clinical laboratory parameters, vital signs, 12-lead ECG results, Lansky play scores or Karnofsky performance scores, physical examination, height, tooth formation abnormalities, closure of proximal tibial plates, Tanner Stage, and left ventricular ejection fraction (LVEF). Study Day 1 for all safety analyses will be defined as the date of the first dose of study drug.

6.6.1 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 as appropriate.

6.6.1.1 Extent of Exposure of Study Drug

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 drug(s). 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 for lenvatinib 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. However, the dosing end date/dose for chemotherapy (ifosfamide and etoposide) will be imputed as the last dose date/dose recorded in the database prior to or on the analysis cutoff date.

The duration of each individual drug in days will be calculated (as Date of the last non-zero dose – Date of the first non-zero dose + 1) 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 any of the following: lenvatinib 14 mg/m^2, ifosfamide 3000 mg/m^2, and etoposide 100 mg/m^2.

6.6.1.2 Study Drug Administration

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/m^2/day) will be calculated as the total dose (mg) received during the study divided by BSA on Cycle 1 Day 1 and 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 planned dose level.

6.6.1.3 Study Drug Dose Modifications

DOSE REDUCTIONS

Dose reduction refers to a situation where a dose level was reduced from the previous dose level without going back. Dose reduction could apply to any of the study drugs.

DOSE INTERRUPTIONS

Dose interruption refers to a situation where a subject had a planned temporary break from taking the study drug for a short period (ie, interruption period). Dose interruptions will only be determined for lenvatinib.

Dose interruption only refers to the scenario where the dose levels or dosing frequencies before and after interruption period (defined as the period with dose=0) are the same. For example: 14 mg/m^2 lenvatinib followed by 0 mg and followed by 14 mg/m^2 lenvatinib. 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. An example of dose reduction instead of dose interruption:

- 14 mg/m^2 lenvatinib followed by 0 mg followed by 11.2 mg/m^2 lenvatinib.

Please note, for the scenario 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 for each individual study drug and with dose interruptions for lenvatinib will be summarized. The 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 lenvatinib for those subjects who had at least one dose reduction during the treatment period.

6.6.2 Adverse Events

Adverse events will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE), Version 5.0. The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the MedDRA Version 25.0. Adverse events will be coded to the MedDRA lower level term (LLT) closest to the verbatim term. The linked MedDRA PT and primary SOC are also captured in the database.

A TEAE is defined as an AE that emerges during treatment (and within 30 days of the last study treatment), having been absent at pretreatment (Baseline) or

- Re-emerges during treatment, having been present at pretreatment (Baseline) but stopped before treatment,

  or

  Worsens in severity during treatment relative to the pretreatment state when the AE is continuous.

Only those AEs that are treatment-emergent will be included in summary tables by treatment group. Exposure-adjusted analyses will be presented for selected AE summaries. For subjects in Arm B who received optional lenvatinib only or lenvatinib plus chemotherapy, adverse events occurring on or after the date of the crossover will be summarized separately by PT in order of decreasing frequency. All AEs, treatment-emergent or otherwise, will be presented in subject data listings.

6.6.2.1 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
- Treatment-emergent SAEs,
- Treatment-related TEAEs,
- TEAEs of Grade 3 or higher
- TEAEs leading to treatment discontinuation (discontinuation in each study arm and discontinuation of all 3 study drugs individually in each treatment arm)
- TEAEs leading to dose reductions
- TEAEs leading to dose interruptions
- TEAEs with fatal outcomes

6.6.2.2 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

6.6.2.3 Treatment-related TEAEs

The following treatment-related TEAEs will be summarized:

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

6.6.2.4 Treatment-emergent SAEs

The incidence of treatment-emergent SAEs will be summarized as below:

- Treatment-emergent SAEs by SOC and PT
- Treatment-related treatment-emergent SAEs by SOC and PT
- Treatment-emergent SAEs by decreasing frequency of PT
- Most common (eg, ≥5%) treatment-emergent SAEs by SOC and PT

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

6.6.2.5 TEAEs Leading to Treatment Discontinuation

The TEAEs leading to treatment discontinuation (discontinuation in each study arm and discontinuation of all 3 study drugs individually in each treatment arm) will be summarized by SOC and PT.

A listing of TEAEs leading to any study drug discontinuation will be provided.

6.6.2.6 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 TEAEs leading to dose reductions and/or dose interruptions of any study drug will be provided.

6.6.2.7 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 common (eg, ≥5%) Grade 3 or higher TEAEs by SOC and PT

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

6.6.2.8 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.

6.6.2.9 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.

6.6.2.10 Adverse Events of Clinical Interest

Clinically significant adverse events (CSAEs) for lenvatinib have been identified based on a thorough review of safety data from existing clinical and pharmacovigilance databases.

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

CSAEs will be summarized by CTCAE grade.

The subjects with CSAEs leading to dose reductions, dose interruptions, and treatment discontinuations will also be provided.

6.6.3 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.

On-treatment laboratory tests will be defined as the laboratory tests conducted from the start of treatment to no more than 30 days after the last dose of study treatment.

6.6.3.1 Hematology and Clinical Chemistry

Laboratory parameters will be graded based on CTCAE Version 5.0. In the summary of laboratory parameters by CTCAE grade, for parameters with CTCAE grading in both high and low directions (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) count, white blood cell (WBC) count, absolute neutrophil count, 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, glucose, lactate dehydrogenase (LDH), amylase, lipase, total protein

Descriptive statistics for values and changes from baseline at each scheduled visit for hematology and chemistry laboratory parameters will be provided.

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

6.6.3.2 Thyroid-Stimulating Hormone

Thyroid-stimulating hormone values will be summarized by category according to standardized reference ranges.

6.6.3.3 Urinalysis

The shift of worst postbaseline proteinuria from baseline will be summarized. All urinalysis parameters (RBCs/high-power-field [HPF], blood, and protein [dipstick], and 24-hour urine protein) will be presented in by-subject listings.

6.6.4 Vital Signs

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

Shifts from baseline will be presented overall and by subgroups of subjects <18 and ≥18. In addition, percentiles for values of height and blood pressure (only for subjects <18 years old [see Protocol Appendices 5 and 6]) will be summarized as a shift from baseline to end of treatment by category for height: 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles; and as a shift from baseline to worst (of SBP or DBP) postbaseline measurement by category: <90th, 90th to 95th, 95th to ≤99th, and >99th percentiles.

6.6.5 Other Safety Analyses

6.6.5.1 Electrocardiograms

Descriptive statistics for ECG parameters (HR, PR, QRS, QT, QTcB, QTcF and RR) and changes from baseline at each scheduled visit will be presented.

Shift tables of worst postbaseline values from baseline will be presented for ECG findings (categorized as normal; abnormal, not clinically significant; and abnormal, clinically significant).

Clinically abnormal ECG results in QTc (QTc Fridericia and QTc Bazett) will be categorized using the following analyses.

Absolute QTc interval prolongation will be summarized for the number (percentage) of subjects with at least 1 postbaseline ECG result during treatment:

- QTc interval ≤450 msec
- QTc interval >450 - 480 msec
- QTc interval >480 - 500 msec
- QTc interval >500 msec

Change from baseline in QTc interval will be summarized for the number (percentage) of subjects with at least one Baseline and postbaseline ECG result during treatment:

- QTc interval increases from baseline >0 - 30 msec
- QTc interval increases from baseline >30 - 60 msec
- QTc interval increases from baseline >60 msec

Subject listings will also be provided.

6.6.5.2 Left Ventricular Ejection Fraction

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

Categorical summaries will also be presented as follows:

- Percentage reduction from baseline (> 0 to ≤10%, >10 to ≤15%, >15%)
- Occurrence of subjects having a lowest postbaseline value of <50%.

A subject data listing will be provided.

6.6.5.3 Lansky Play Scores or Karnofsky Performance Status Scores

Lansky play scores or KPS scores will be presented as a figure showing results over time. A subject listing of Lansky play scores or KPS scores will be provided.

6.6.5.4 Proximal Tibial Growth Plates

Radiographic findings of proximal tibial growth plates will be summarized at baseline and end of treatment.

A subject listing will be provided.

6.6.5.5 Tanner Stage

Tanner Stage will be summarized across time by visit, and as a shift from baseline to end of treatment.

A subject listing will be provided.

6.6.5.6 Dental Examination

Tooth formation abnormalities will be summarized and listed.

6.6.5.7 New Anticancer Therapies During Follow-up Period

The new anticancer medications and procedures during Follow-up Period will be reported and listed.

6.7 Other Analyses

6.7.1 Palatability

Each question in the Palatability and Acceptability Questionnaire will be summarized and listed.

6.7.2 Health-Related Quality of Life

Health-related quality of life will be collected at baseline (Day -1 or at Cycle 1 Day 1 prior to treatment) and C2D1, C3D1, Week 18, C8D1, C18 D1 and at the Off-Treatment Visit. Score change from baseline in PedsQL at each analysis timepoint will be analyzed. The primary timepoint for assessment is at Week 18 for all PedsQL endpoints.

Descriptive statistics will be presented for all PedsQL endpoints at each analysis time period by treatment arm. Baseline is defined as the later value of Day -1 or at Cycle 1 Day 1 prior to treatment.

A detailed HRQoL analysis plan will be provided in a separate document and the results will be provided in a stand-alone report.

6.8 Exploratory Analyses

Other exploratory analyses may be conducted as appropriate.

7 INTERIM ANALYSES

No interim analysis is planned for this study.

8 CHANGES IN THE PLANNED ANALYSES

Based on the blinded review of the data, the number of subjects in the level of the stratum "time to first relapse/refractory disease ≥18 months" is only about 11 subjects out of the ITT total size of 81 subjects. To ensure a sufficient number of participants in each stratum within each treatment arm, "time to first relapse/refractory disease**"** stratification factor was removed from the analyses. Only the stratification factor "age at randomization" will be included in any stratified analyses.

For DCR, the SD minimum duration period is changed from 7 weeks to 5 weeks after randomization. The change was made due to the tumor assessment schedule of every 6 weeks until Week 18.

9 DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

9.1 Visit Windows

Visit windows will be defined to be within upper and lower bounds of 3 days of the scheduled visit, according to the protocol, which states that efforts should be made to conduct study visits (and safety assessments) on the day scheduled (± 1 day for Cycle 1 and within 3 days for Cycle 2 onward).

Tumor assessments should be performed at screening and every 6±1weeks after the date of randomization or sooner if clinically indicated until Week 18±1 week. Then every 9±1weeks until Week 54±1 week, and thereafter, every 12±2 weeks until documentation of PD. If a subject discontinues from study treatment without disease progression, tumor assessments should continue to be performed using this schedule.

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 scheduled and unscheduled visits.

9.2 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 any study drug (ie, lenvatinib, ifosfamide, or etoposide).

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.

9.3 Imputation of Missing Data

Unless specified otherwise, no data imputation will be applied for missing safety and efficacy evaluations. For analysis and reporting purposes, 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 specifications with more details.

9.4 Variable Derivations

9.4.1 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 osteosarcoma diagnosis to date of randomization (months)
- Time since last disease progression to date of randomization (months)
- Time from end of most recent therapy to date of randomization (months)
- Time from previous surgery to date of randomization (months)
- Time from most recent radiotherapy to date of randomization (months)

9.5 Pharmacokinetics/Pharmacodynamics Data Handling

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

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.

10 PROGRAMMING SPECIFICATIONS

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

11 STATISTICAL SOFTWARE

Statistical programming and analyses will be performed using SAS® (SAS Institute, Inc., Cary, NC, USA), version 9.0 or higher, and/or other validated statistical software as required.

12 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 detail.

13 REFERENCES

1. Brookmeyer R, Crowley JJ. A confidence interval for the median survival time. Biometrics. 1982:38:29-41.
2. Clopper CJ Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika. 1934;26:404–13.
3. Common Terminology Criteria for Adverse Events (CTCAE). Version 5.0. United States Department of Health and Human Services, National Institutes of Health. National Cancer Institute, Washington, DC, USA, November 27, 2017.
4. Davis LE, Bolejack V, Ryan CW, Ganjoo KN, Loggers ET, Chawla S, et al. Randomized double-blind phase II study of regorafenib in patients with metastatic osteosarcoma. J Clin Oncol. 2019 Jun 1;37(16):1424-31.
5. 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.
5. Greenwood, M. The natural duration of cancer. Reports on Public Health and Medical Subjects. 1926:33:1-26. Her Majesty's Stationery Office, London.
6. Food and Drug Administration Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, May 2007 [internet; cited 3 March 2011]. Available from: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf.
7. Grignani G, Palmerini E, Dileo P, Asaftei SD, D'Ambrosio L, Pignochino Y, et al. A phase II trial of sorafenib in relapsed and unresectable high-grade osteosarcoma after failure of standard multimodal therapy: an Italian Sarcoma Group study. Ann Oncol. 2012;23(2):508-16.
8. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd Edition. John Wiley & Sons, Inc. New York. 2002.
9. Kaplan E, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53:457-81.
10. Palmerini E, Jones RL, Marchesi E, Paioli A, Cesari M, Longhi A, et al. Gemcitabine and docetaxel in relapsed and unresectable high-grade osteosarcoma abd spindle cell sarcoma of bone. BioMed Centr Cancer. 2016;16:280.
11. Guideline on the evaluation of anticancer medicinal products in man - methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials, 13 December 2012. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf.

14 APPENDICES

Not applicable for this report.

SIGNATURE PAGE

Author:

PPD

6/17/2022

Date

Oncology Business Group, Eisai Inc.

Approval:

PPD

6/20/2022

Date

Oncology Business Group, Eisai Inc.

PPD

6/20/2022

Date

Oncology Business Group, Eisai Ltd.