J2G-MC-JZJB Statistical Analysis Plan Version 3

LIBRETTO-531:A Multicenter, Randomized, Open-Label, Phase 3 Trial Comparing Selpercatinib to Physicians Choice of Cabozantinib or Vandetanib in Patients with Progressive, Advanced, Kinase Inhibitor Naïve, RET-Mutant Medullary Thyroid Cancer

NCT04211337

Approval Date: 16-Dec-2021

1. Statistical Analysis Plan for Clinical Study J2G-MC-JZJB

LIBRETTO-531: A Multicenter, Randomized, Open-Label, Phase 3 Trial Comparing Selpercatinib to Physicians Choice of Cabozantinib or Vandetanib in Patients with Progressive, Advanced, Kinase Inhibitor Naïve, RET-Mutant Medullary Thyroid Cancer

LY3527723 (Selpercatinib; LOXO-292)
RET-Mutant Medullary Thyroid Cancer

This is a phase 3, multicenter, randomized, open-label trial comparing selpercatinib (LOXO-292) to Physicians Choice of Cabozantinib or Vandetanib in Patients with Progressive, Advanced, Kinase Inhibitor Naïve, RET-Mutant Medullary Thyroid Cancer.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol J2G-MC-JZJB
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 18-Dec-2019
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 22-Feb-2021
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on date provided below:

Approval Date: 16-Dec-2021 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical analysis plan (SAP) Version 1 was approved prior to the first patient visit.

Version 2 was approved prior to the first permanent data transfer.

Version 3 was approved prior to the first interim efficacy analysis to update the components listed below in response to,

- Regulatory feedback:
    - Added progression free survival (PFS), as a primary endpoint and removed it as a secondary endpoint
    - Removed treatment failure free survival (TFFS), as a primary endpoint and added it as a secondary endpoint
- New disclosure of external data: The study has been updated to an adaptive design to allow sample size re-estimation based on interim PFS results

Section # and Name	Description of Change	Brief Rationale
4.1 Objectives and Endpoints	Added PFS as a primary endpoint and removed it as a secondary endpoint Removed TFFS as a primary endpoint and added it as a secondary endpoint.	Based on regulatory agency feedback
4.1 Objectives and Endpoints	Added safety assessment of selpercatinib in patients assigned to Arm B who crossover to selpercatinib after progression in exploratory objectives and endpoints	Based on local clinical feedback
4.2 Study Design	Revised the patient enrollment data and updated the section with changes in primary endpoint and key secondary endpoint.	Change to adaptive study design
5 Statistical Hypotheses	Removed TFFS and added PFS evaluation for the treatment of patients	Based on regulatory agency feedback
6 Sample Size Determination	Revised the section to include adaptive study design and sample size re-estimation based on PFS.	Change to adaptive study design
7 Populations for Analyses	Updated the description of crossover population	Clarification
7 Populations for Analyses	Updated the description of tolerability evaluable population	Change according to the adaptive study design
8.1 General Considerations	Added definition of the core study period	Clarification
8.1.1 Definitions	Added definition of crossover baseline measurement	Clarification based on local clinical feedback
8.3 Primary Endpoint Analyses	Added PFS as the primary endpoint. Removed TFFS as the primary endpoint. Added description of estimand for PFS analysis Updated the main analytical approach for PFS	Based on regulatory agency feedback. Change to adaptive design method
8.4.1.1 TFFS by BICR	Added TFFS as a key secondary endpoint. Added description of estimand for TFFS analysis Updated the main analytical approach for TFFS	Based on regulatory agency feedback. Change to adaptive design method
8.4.1.2 Comparative Tolerability	Modified the analysis population and sample size	Change according to the adaptive study design

Section # and Name	Description of Change	Brief Rationale
8.4.2 Supportive Secondary Endpoints	Added the time requirement of confirmed response. Added definition of disease control rate	Clarification
8.4.2 Supportive Secondary Endpoints	Updated OS analysis plan	Change according to the adaptive study design
8.4.2 Supportive Secondary Endpoints	Modified the definition of treatment-emergent adverse events	Clarification
8.5 Tertiary/Exploratory Endpoint Analyses	Added efficacy and safety analyses after crossover	Based on local clinical feedback
8.8 Interim Analyses	Updated the trigger for interim analyses and guidance for the study results categorization at interim analysis	Change to adaptive study design
8.8.1 Maintaining the Trial Integrity	Added the plan to main the trial integrity for adaptive design	Based on regulatory agency feedback
8.8.2 Safety Update Report	Added analyses for periodic safety update report and Japan periodic safety review	Based on regulatory agency feedback

4. Introduction

4.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
• To compare PFS of patients with progressive, advanced, kinase inhibitor naïve, RET-mutant MTC treated with selpercatinib versus cabozantinib or vandetanib	• PFS by BICR
Secondary	
• To compare other efficacy outcomes, based on RECIST 1.1 criteria, observed in patients with progressive, advanced, kinase inhibitor naïve, RET-mutant MTC treated with selpercatinib versus cabozantinib or vandetanib	• TFFS by BICR • TFFS by investigator • PFS by investigator • ORR by investigator and BICR • DOR by investigator and BICR • OS • PFS2 by investigator
• To evaluate the safety and tolerability of selpercatinib compared to cabozantinib or vandetanib	• Safety per CTCAE v5.0 (including but not limited to): incidence and severity of TEAEs, SAEs, deaths, and clinical laboratory abnormalities
• To compare the tolerability of selpercatinib versus cabozantinib or vandetanib	• Proportion of time with high side effect bother based on FACT-GP5
• To assess/evaluate the performance of local RET laboratory tests compared to a single, central test	• RET mutation status
• To assess the PK of selpercatinib in the patient population	• Predose plasma concentrations at Day 8 of Cycle 1 and at Day 1 of Cycles 2 through 6
Tertiary/Exploratory	
• To compare the calcitonin and CEA response rate of patients with progressive, advanced, kinase inhibitor naïve, RET-mutant MTC treated with selpercatinib versus cabozantinib or vandetanib	• Calcitonin response rate • CEA response rate
• To compare the PROs of disease-related symptoms, symptomatic adverse events and overall side effect burden, physical function, and HRQoL of patients with progressive, advanced, kinase inhibitor naïve, RET-mutant MTC treated with selpercatinib versus cabozantinib or vandetanib	• Bristol Stool Form Scale and bowel movement frequency • Physical Function (EORTC IL19) • HRQoL (EORTC QLQ-C30-PF) • Health Utilities (EQ-5D-5L) • Worst Pain NRS • PRO-CTCAE
• To assess the efficacy and safety of selpercatinib in patients assigned to Arm B who crossover to selpercatinib after progression	• PFS after crossover • Safety per CTCAE v5.0 after crossover (including but not limited to): incidence and severity of TEAEs, SAEs, deaths, and clinical laboratory abnormalities.
• To compare the TTNT of patients with progressive, advanced, kinase inhibitor naïve, RET-mutant MTC treated with selpercatinib versus cabozantinib or vandetanib	• TTNT

Objectives	Endpoints
• To compare RET mutation status in tumor and genomic DNA samples	• Biomarker analyses
• To assess the relationship between biomarkers and clinical outcomes	• Biomarkers assessed from blood or tissue samples unless precluded by local regulations • Clinical outcomes data

Abbreviations: BICR = blinded independent review committee; CEA = carcinoembryonic antigen; CTCAE = Common Terminology of Criteria for Adverse Events; DNA = deoxyribonucleic acid; DOR = duration of response; FACT-GP5 = Functional Assessment of Cancer Therapy-Side Effects; EORTC QLQ-30-C30-PF = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 Version 3.0; EORTC IL19 = European Organisation for Research and Treatment of Cancer, item library 19; EQ-5D-5L = 5-level-EuroQol; HRQoL = health-related quality of life; MTC = medullary thyroid cancer; NRS = numeric rating scale; ORR = overall response rate; OS = overall survival; PFS = progression free survival; PFS2= progression-free survival 2; PK = pharmacokinetic; PRO = patient reported outcome; RECIST = Response Evaluation Criteria in Solid Tumors; RET = rearranged during transfection; SAE = serious adverse event; TEAE = treatment-emergent adverse event; TFFS = treatment failure-free survival; TTNT = time to initiation of new anticancer therapy.

4.2. Study Design

This is a global, multicenter, randomized (2:1), open-label, Phase 3 study comparing selpercatinib (treatment Arm A) to physicians' choice of cabozantinib or vandetanib (treatment Arm B) in patients with progressive, advanced, kinase inhibitor naïve, rearranged during transfection (RET)-mutant medullary thyroid cancer (MTC).

Patients will be randomized in a 2:1 ratio to selpercatinib or physician's choice of cabozantinib or vandetanib.

Patients will be stratified based on:

- RET mutation: M918T vs. other
- Intended treatment if randomized to control arm: cabozantinib vs. vandetanib

Approximately 250 patients will be initially enrolled in the study. A sample size re-estimation based on comparative data will be conducted at the interim efficacy analysis. The total number of patients could be increased from the initially planned 250 up to a maximum of approximately 400 depending on the results of the interim efficacy analysis. The total study duration will be capped at 6 years from the first patient visit.

Patients with histologically confirmed, unresectable, locally advanced, or metastatic MTC who have not received previous treatment with a kinase inhibitor are eligible. Patients are required to have radiologic progressive disease (PD) per Response Evaluation Criteria in Solid Tumors, version 1.1 (RECIST 1.1, Eisenhauer et al. 2009) at screening compared with an image obtained within the prior 14 months. Patients are also required to have a documented RET mutation in tumor or germline deoxyribonucleic acid (DNA) identified through molecular assays.

After confirmation of eligibility, patients will be randomly assigned in a 2:1 ratio to:

- Arm A: selpercatinib at a starting dose of 160 twice a day or
- Arm B: cabozantinib at a starting dose of 140 mg once daily, or vandetanib at a starting dose of 300 mg once daily. Patients cannot switch from cabozantinib to vandetanib or from vandetanib to cabozantinib during the study.

Treatment will continue until disease progression, unacceptable toxicity, or death. The treatment decision will be made by investigator assessment. Patients with disease progression per investigator assessment may continue treatment while awaiting blinded independent central review (BICR) confirmation of progression.

Patients discontinuing treatment for any reason other than death, lost to follow-up, or withdrawal of consent will enter the survival follow-up period and be followed every 3 months for the development of radiographic disease progression (if not already occurred) and initiation of subsequent anticancer therapies until death, lost to follow-up, or withdrawal of consent (whichever comes first).

Patients who discontinue treatment and who have radiographic disease progression that is confirmed by BICR and were randomized to cabozantinib or vandetanib may be eligible for crossover to selpercatinib if they meet the eligibility criteria for crossover.

A detailed description of the study design is contained in the protocol.

5. Statistical Hypotheses

Treatment of patients with progressive, advanced, kinase inhibitor-naïve, RET-mutant MTC with selpercatinib in the first line setting will provide a clinically meaningful increase in PFS over treatment with cabozantinib/vandetanib.

6. Sample Size Determination

An adaptive design with a sample size re-estimation based on comparative data at the interim efficacy analysis was selected to mitigate the uncertainties of the true treatment effect by allowing the study to adjust to information that is currently not available. There is significant uncertainty as to the performance of both the treatment Arm A and Arm B. Specifically:

- Monitoring of MKI-naïve patients participating in LIBRETTO-001 provides some insight into the potential outcomes of patients treated with selpercatinib on Arm A. Reliable prediction of median PFS for the Arm A based on the LIBRETTO-001 data is challenging due to the very immature data with a high censoring rate at the time of SAP version 3 approval.
- Arm B uses two drugs with very large differences in reported median PFS time. It is unclear if the differences are due to differences in the treated population, inherent differences in drug effects, or both. However, these factors all contribute to the uncertainty of estimating median PFS for Arm B.
    - In the EXAM study, the median PFS was 11 months for cabozantinib-treated patients with documented progression within 14 months of treatment enrollment (irrespective of the presence of a *RET* mutation) (Elisei et al. 2013, Schlumberger et al. 2017), and was longer (13.9 months) for the subset of patients with cancers harboring *RET* M918T mutations (the most common *RET* mutation) (Sherman et al. 2016).
    - In the ZETA study, the median PFS was 30.5 months for vandetanib-treated patients with newly diagnosed advanced MTC and patients with progression (Wells et al. 2012). A post hoc analysis provided median PFS results for two subgroups of patients that are similar to those enrolling in Study JZJB based on the presence of progressive disease within 12 months of enrollment. For those with both disease progression and symptoms in the prior 12 months, the median PFS was 21.4 months and for those with disease progression only in the prior 12 months, median PFS was not reached at a median follow-up of 95 months (Kreissl et al. 2020).

An initial assumption of 74 PFS events provides approximately 80% power to detect a hazard ratio (HR) of 0.5 using the logrank test and a 1-sided type I error of 2.5%. Given the historical data outlined above, a median PFS of 25 months is assumed for Arm B. HR of 0.5 corresponds to an improvement in median PFS from 25 months for Arm B to 50 months for Arm A. Approximately 250 patients will be enrolled based on these initial assumptions.

The sample size re-estimation will be based on PFS, so the sample size adjustment has the purpose of modifying the number of events. The number of patients will be adjusted to ensure the achievement of the required total number of events in an expeditious manner.

A re-estimation of the number of PFS events will be conducted only once, during the pre-specified interim efficacy analysis, based on the unblinded comparative results observed at this analysis. Based on prespecified criteria described in the Adaptive Design Charter the following scenarios are possible:

- the study will be declared positive due to overwhelming efficacy
- the study will continue without change to the final analysis (ie, the re-estimated total number of events will be equal to the initial planned total number of events)
- the re-estimated total number of events required for the final analysis will be determined and the study will continue to the final analysis.

The re-estimated total number of events could be increased from 74 to a maximum of 284 to maintain the conditional power (conditional probability of a statistically significant treatment effect at the end of the trial) at a prespecified level. A maximum of 284 PFS events was selected based on the original protocol under a fixed study design to provide approximately 80% power to detect a HR of 0.7. The total number of patients could be increased from approximately 250 up to approximately 400. The total study duration will be capped at 6 years from the first patient visit regardless of the actual number of events observed. Details of adaptation decision rules are described in Section 8.8 and the Adaptive Design Charter.

7. Analysis Sets

For purposes of analysis, the following populations are defined:

Population	Description
Entered	All participants who sign an informed consent
Intention to treat (ITT)/ Enrolled	All randomized patients, even if a patient does not take the assigned treatment, does not receive the correct treatment, or otherwise does not follow the protocol. Patients will be analyzed according to the treatment arm they were assigned to regardless of what actual treatment they receive
Per-protocol (PP)	All randomized patients (ITT population) who do not have important protocol deviations (IPDs) that could potentially affect the efficacy conclusions of the study
Evaluable	Defined in the following specific subsections if applicable
Crossover	A subpopulation of patients included in the ITT population who were randomly assigned to Arm B, crossed over and took at least 1 dose of selpercatinib
Safety	All randomized patients who take at least 1 dose (including a partial dose) of study treatment. Analysis of safety data will be based on the actual treatment a patient received on the first study treatment administration regardless of which treatment they were randomized to receive (“as treated”)
Tolerability Evaluable	All patients who received the first dose of study treatment prior to the interim efficacy analysis and at least 6 months prior to the data cutoff date. Analysis of tolerability will be based on the actual treatment a patient received on the first study treatment administration regardless of which treatment they were randomized to receive (“as treated”)

A patient listing of analysis population details will be provided. This listing will be presented by the treatment arm and will include: investigator site, patient identifier, inclusion/exclusion flag for each population, and reason for exclusion from each population. All patients screened will appear on this listing.

8. Statistical Analyses

8.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, unless otherwise stated, and all confidence intervals (CIs) will be given at a 2-sided 95% level, unless otherwise stated. Endpoints and analyses are defined for the core study period, ie study period with randomized treatment before crossover, unless otherwise specified. Statistical analysis will be performed using SAS software (SAS, version 9.1.2 or higher).

Continuous variables will be summarized using descriptive statistics (ie, number of patients, mean, median, standard deviation, minimum, and maximum). Categorical variables will be summarized by frequency and their corresponding percentage.

Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the clinical study report. Additional exploratory analyses of the data will be conducted as deemed appropriate.

8.1.1. Definitions

Definitions of efficacy, safety, and patient-reported outcome (PRO) analysis variables are listed in respective sections of the SAP. Other variables are listed below alphabetically:

- Age (years): (informed consent date - date of birth + 1)/365.25; birth month and day are imputed to be 01 July, because only the birth year is collected through electronic case report form (eCRF).
- Baseline Measurement: unless otherwise specified, the last nonmissing measurement prior to the first dose of study drug for safety and tolerability analyses, and the last nonmissing measurement prior to randomization for baseline and efficacy analyses.
- Crossover Baseline Measurement: unless otherwise specified, the last nonmissing measurement prior to the first dose of selpercatinib during crossover period for safety, tolerability, and efficacy analyses.
- Duration: duration is calculated as:
    - Duration (days): (end date – start date + 1)
    - Duration (weeks): (end date – start date + 1)/7
    - Duration (months): (end date – start date + 1)/30.4375
      (days in months = (1/12) * average number of days in a year)
    - Duration (years): (end date – start date + 1)/365.25
- Duration of disease: (randomization date - diagnosis of cancer date + 1).
- Study Day (safety and tolerability analyses): study day is calculated as assessment date – first dose date + 1 day if the assessment is done on or after the first dose day. If the assessment is done prior to the first dose day, the study day will be calculated as assessment date – first dose date. Date of first dose is defined as Study Day 1.

- Study Day (baseline and efficacy analyses): study day is calculated as assessment date – randomization date + 1 day if the assessment is done on or after randomization. If the assessment is done prior to randomization, the study day will be calculated as the assessment date – randomization date. Date of randomization is defined as Study Day 1.
- Time-to-Event: the event or censoring time (days) is calculated as the date of event/censoring – randomization date + 1.

8.1.2. Handling of Dropouts or Missing Data

All analyses and descriptive summaries will be based on the observed data. Unless otherwise specified, missing data will not be imputed or "carried forward". Rules for handling dropouts or missing data are listed by type of analysis alphabetically.

- Adverse event (AE) or concomitant therapy:
    - The missing day of onset of an AE or start date of a concurrent therapy will be set to:
        - first day of the month that the event occurred, if the onset yyyy-mm is after the yyyy-mm of first study treatment;
        - the day of the first study treatment, if the onset yyyy-mm is the same as yyyy-mm of the first study treatment; or
        - the date of informed consent, if the onset yyyy-mm is before the yyyy-mm of the first treatment.
    - The missing day of resolution of an AE or end date of a concurrent therapy will be set to:
        - the last day of the month of the occurrence. If the patient died in the same month, then set the imputed date as the death date.
    - If the onset date of an AE or start date of a concurrent therapy is missing both the day and month, the onset date will be set to:
        - 01 January of the year of onset, if the onset year is after the year of the first study treatment;
        - the date of the first treatment, if the onset year is the same as the year of the first study treatment; or
        - the date of informed consent, if the onset year is before the year of the first treatment.
    - If the resolution date of an AE or end date of a concurrent therapy is missing both the day and month, the date will be set to:
        - 31 December of the year of occurrence. If the patient died in the same year, then set the imputed date as the death date.
    - If the date is completely missing, then no imputation will be done and the event will be considered as treatment-emergent with unknown onset date, unless the end date rules out the possibility.
- Diagnosis date, the following conventions will be used for imputing partial dates:
    - If only the day of the month is missing, the 15th of the month will be used to replace the missing day.

  - If both the day and the month are missing, "Jul 1" will be used to replace the missing information.
- General rule for imputing other dates:
  - If only the day is missing, then assign Day 15 of the month, or the date of death if the patient died prior to 15th of the same month to the day.
  - If month is missing, then the date will be set to July first of the year, or the date of death if the patient died prior to July first of the same year.

  However, in all cases, after imputation, check if the imputed date is logically consistent with other relevant date variable(s) and make appropriate correction if necessary.

- Relationship: missing classifications concerning relationship will be considered as related to study medication.
- Time-to-event analysis: all censored data will be accounted for using appropriate statistical methods.

8.2. Participant Disposition

A detailed description of participant disposition will be provided according to the Consolidated Standards of Reporting Trials (CONSORT) publishing requirements, including a summary of the number and percentage of patients entered into the study, enrolled in the study, and treated, as well as number and percentage of patients completing the study, or discontinuing (overall and by reason for discontinuation).

8.2.1. Protocol Deviations

Important protocol deviations (IPDs) that could potentially compromise the data integrity and patients' safety will be summarized for the intention to treat (ITT) population. These deviations will include deviations that can be identified programmatically and those that can only be identified by the clinical research associate during monitoring. Important protocol deviations are described in the Trial Issue Management Plan (TIMP) within the study Trial Master File.

The list of IPDs that could potentially affect the efficacy conclusions of the study will be defined and documented in the TIMP prior to the final database lock in order to identify patients to be excluded from the per-protocol (PP) population.

8.3. Primary Endpoint Analyses

8.3.1. Research Objective and Question

The primary research objective and question is: What is the difference in PFS time between selpercatinib and cabozantinib/vandetanib as the first line therapy in patients with progressive, advanced, kinase inhibitor-naïve, RET-mutant MTC.

8.3.2. Definition of Endpoint

The estimand for the primary objective is described by the following attributes:

- Population: patients with progressive, advanced, kinase inhibitor-naïve, RET-mutant MTC. Further details can be found in Protocol Section 5 Study Population.
- Endpoint: progression-free survival confirmed by BICR, which is defined as the time from randomization until the occurrence of documented disease progression by the BICR, per RECIST 1.1 criteria, or death from any cause in the absence of BICR-documented PD.
- Treatment condition: the randomized study intervention (selpercatinib or cabozantinib/vandetanib) will be administered twice or once a day in continuous 28-day cycles until disease progression, unacceptable toxicity or other protocol-defined reasons for treatment discontinuation. Further details on study interventions including interventions, concomitant therapy and dose modification can be found in Protocol Section 6 Study Intervention.
- Intercurrent-event strategies (IES) for main analytical approach are listed in Table JZJB.1.
- Population-level summary measure: hazard ratio of PFS in selpercatinib versus cabozantinib/vandetanib estimated using a stratified Cox regression model (Cox 1972)

Table JZJB.1. Intercurrent Events of PFS

Intercurrent Event	IES
Study intervention discontinuation prior to PFS event	"Treatment policy" strategy: regardless of whether or not study intervention discontinuation had occurred
Post study intervention discontinuation anticancer therapy prior to PFS event	"While on treatment" strategy: consider the assessment of endpoint up until the time that post study intervention discontinuation anticancer therapy is taken
Extended time without adequate assessment prior to PFS event	"While on treatment" strategy: consider the assessment of endpoint up until the occurrence of extended time without adequate assessment

Abbreviation: IES = intercurrent-event strategies; PFS = progression-free survival.

Rationale for IES: The interest lies in the treatment effect without the confounding effect of other anticancer therapy or extended time without adequate assessment.

- Study intervention discontinuation due to reasons other than PFS event is handled with treatment policy as it reflects clinical practice. Time from randomization until disease progression or death regardless of study intervention discontinuation will be considered in analysis.
- A post study intervention discontinuation anticancer therapy taken prior to PFS event will confound the treatment effect of selpercatinib in terms of PFS. If the anticancer therapy is taken, future disease progression/death status is not needed. The participant will be censored and only the time prior to the post study intervention discontinuation anticancer therapy will be considered in analysis.

- PFS event observed after an extended time without adequate tumor assessment may have occurred much earlier but is not reported because the scheduled assessment was not done. This inadequate observation may introduce bias to PFS estimates. If extended time without adequate assessment occurs, the participant will be censored and only the time up to the last adequate tumor assessment will be considered in analysis.

8.3.3. Main Analytical Approach

PFS by BICR will be compared between treatment arms using the Cui, Hung, and Wang (CHW) testing procedure (Cui et al. 1999) to control the type I error at an overall 1-sided 2.5% significance level:

- At the interim efficacy analysis, the PFS by BICR will be compared using a conventional stratified logrank test, stratified by the 2 randomization strata based on interactive web-response system (IWRS) data: RET mutation (M918T vs. other) and intended control treatment (cabozantinib vs. vandetanib).
- At the final analysis, the PFS by BICR will be compared using a CHW test as described below. If the re-estimated total number of events is equal to the initial planned total number of events, the CHW test will be reduced to the conventional stratified logrank test.

Table JZJB.2. Defining the Values for Computing the CHW Test Statistic of PFS

Quantity	Input Value or Derivation / Calculation
$z_{PFS,1}$	This is the stratified logrank test statistic comparing PFS between treatment arms at the interim analysis. The interim analysis is triggered after approximately 56 PFS events have occurred (at 75% information fraction).
$z_{PFS,2}$	This is the stratified logrank test statistic comparing PFS between treatment arms at the final analysis.
$e_{PFS,1}$	This is the observed number of PFS events across both treatment arms at the interim analysis. The interim analysis is triggered after approximately 56 PFS events have occurred (at 75% information fraction). Due to the uncertainty in operation, the exact timing of the interim may occur at a number of events slightly smaller or larger than the planned 56 number of PFS events.
$e_{PFS,2}$	This is the number of PFS events across both treatment arms at the final analysis should the study remain at the initial planned total number of events. $e_{PFS,2} = 74$
$e^*_{PFS,2}$	This is the observed number of PFS events across both treatment arms at the final analysis.

Given the quantities defined in Table JZJB.2, the final CHW test statistic for the primary efficacy analysis (at the final analysis after the sample size re-estimation) can be written as a weighted combination of the independent increments comprising the interim stratified logrank test statistic and the post-interim stratified logrank test statistic (Cui et al. 1999):

$$z_{PFS,CHW} = \sqrt{\frac{e_{PFS,1}}{e_{PFS,2}}} z_{PFS,1} + \sqrt{\frac{e_{PFS,2} - e_{PFS,1}}{e_{PFS,2}}} \left\{ \frac{\sqrt{e^*_{PFS,2}} z_{PFS,2} - \sqrt{e_{PFS,1}} z_{PFS,1}}{\sqrt{e^*_{PFS,2} - e_{PFS,1}}} \right\}$$

As pointed out by Cui, Hung, and Wang (Cui et al. 1999), for group sequential test based on the repeated significance test that can be asymptotically expressed as a Brownian motion process, eg logrank test, with independent increment property, if the conventional logrank statistic is replaced by the CHW statistic while using the original rejection boundary for the conventional logrank statistic, the new CHW test will have the total type I error preserved at the specified level. Therefore, to determine the statistical significance at the final analysis, the $z_{PFS,CHW}$ will be compared to the Z scale critical boundary -1.960 given in Section 8.8 Interim Analysis.

In addition, unadjusted HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972), stratified by randomization strata. Progression free survival curves, medians, and PFS rates at various time points with 95% CI for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Table JZJB.3 defines censoring rules to be applied to the BICR PFS main analysis.

Table JZJB.3. PFS Censoring Scheme

Situation	Event/Censor	Date of Event or Censor
Tumor progression[a] by BICR or death	Event	Earliest date of PD by BICR or death
No tumor progression by BICR and no death	Censored	Date of last adequate tumor assessment[b], per RECIST 1.1 criteria, or date of randomization (whichever is later)
Unless		
No baseline radiologic tumor assessment available	Censored	Date of randomization
No adequate postbaseline tumor assessment available and death reported after 2 scan intervals[c] following randomization	Censored	Date of randomization
New systemic anticancer therapy	Censored	Date of adequate tumor assessment, per RECIST 1.1 criteria, prior to start of new therapy or date of randomization (whichever is later)
Tumor progression by BICR or death documented immediately after 2 or more missing scan intervals following last adequate tumor assessment or randomization (whichever is later)	Censored	Date of last adequate tumor assessment, per RECIST 1.1 criteria, or date of randomization (whichever is later)

Abbreviations: BICR = blinded independent central review; CR = complete response; PD = progressive disease; PFS = progression-free survival; PR = partial response; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors, version 1.1, SD = stable disease.

[a] Symptomatic deterioration (that is, symptomatic progression that is not radiologically confirmed per RECIST 1.1 criteria) will not be considered as tumor progression.

[b] Adequate tumor assessment per RECIST 1.1 criteria refers to an assessment with 1 of the following responses: CR, PR, SD, or PD.

[c] The 2-scan interval is counted from the date of last adequate tumor assessment to the date of next 2 scheduled tumor assessments plus 14 days (adjusted by tumor assessment window).

- If there are multiple dates associated with 1 assessment, the assessment date will be set to the first date when the overall response is PD and the last date otherwise.

8.3.4. Sensitivity Analyses

Multiple sensitivity analyses for the PFS by BICR will be conducted as defined below:

- discounting the event of PD if pleural effusion is the only reason of PD per BICR and the patient is experiencing chylous effusion at the time of PD (see Sensitivity Analysis [SA] 6 Table JZJB.4);
- ignoring new systemic anticancer therapy (see SA2 Table JZJB.4) according to the EMA guidelines of censoring scheme (EMA 2012);
- using other rules for censoring (as defined by Table JZJB.4);
- using an unstratified logrank test and an unstratified Cox regression model;
- using stratification factors based on the eCRF data;
- repeating the main PFS by BICR analysis for the PP population.

Table JZJB.4. PFS Censoring Scheme - Sensitivity Analyses

Definition	Situation	Event/ Censor	Date of Event or Censor
SA1: Ignoring absence of adequate postbaseline tumor assessment	No adequate postbaseline tumor assessment available and death reported after 2 scan intervals following randomization	Event	Death
SA2: Ignoring new systemic anticancer therapy	New systemic anticancer therapy and:		
	1. No tumor progression by BICR and no death	1. Censored	1. Date of last adequate tumor assessment, per RECIST 1.1 criteria, or date of randomization (whichever is later)
	2. Tumor progression by BICR or death after start of new therapy	2. Event	2. Earliest date of PD or death
SA3: Considering events right after the new systemic anticancer therapy starts	Tumor progression by BICR or death within next 14 days from the start of new systemic anticancer therapy	Event	Earliest date of PD by BICR or death
SA4: Ignoring missing tumor assessments	Tumor progression by BICR or death documented after 2 or more missing scan intervals following last adequate tumor assessment or randomization (whichever is later)	Event	Earliest date of PD by BICR or death
SA5: Considering lost to follow up as tumor progression	Patient lost to follow up and no tumor progression by BICR and no death	Event	Date of next scheduled tumor assessment at or after patient was lost to follow up
SA6: Discounting PD confounded by chylous effusion	Pleural effusion is the only reason of PD per BICR and the patient is experiencing chylous effusion at the time of PD	Censor	Date of last adequate tumor assessment, per RECIST 1.1 criteria

Abbreviations: BICR = blinded independent central review; PD = progressive disease; PFS = progression-free survival; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors, version 1.1; SA = sensitivity analysis.

8.4. Secondary Endpoints Analyses

8.4.1. Key Secondary Endpoints

8.4.1.1. TFFS by BICR

8.4.1.1.1. Research Objective and Question

A secondary research objective and question is: What is the difference in TFFS time between selpercatinib and cabozantinib/vandetanib as the first line therapy in patients with progressive, advanced, kinase inhibitor-naïve, RET-mutant MTC.

8.4.1.1.2. Definition of Endpoint

The estimand for this secondary objective is described by the following attributes:

- Population: patients with progressive, advanced, kinase inhibitor-naïve, RET-mutant MTC. Further details can be found in Protocol Section 5 Study Population.
- Endpoint: treatment failure free survival by BICR, which is defined as the time from randomization to the first occurrence of:
    - documented radiographic disease progression per RECIST 1.1 as assessed by BICR; or
    - unacceptable toxicity leading to treatment discontinuation as assessed by the investigator. To qualify as an event, the toxicity must be from an intolerable AE (defined as any study drug-related AE that meets protocol guidance for treatment discontinuation, with the exception of alopecia); or
    - death (due to any cause).

  An independent review committee will review blinded data to determine whether an AE leading to treatment discontinuation meets protocol guidance and thus should be considered as a TFFS event. Independent review will occur retrospectively and will be used in analysis to identify the TFFS events.

- Treatment condition: the randomized study intervention (selpercatinib or cabozantinib/vandetanib) will be administered twice or once a day in continuous 28-day cycles until disease progression, unacceptable toxicity or other protocol-defined reasons for treatment discontinuation. Further details on study interventions including interventions, concomitant therapy and dose modification can be found in Protocol Section 6 Study Intervention.
- Intercurrent-event strategies for main analytical approach are listed in Table JZJB.5.
- Population-level summary measure: hazard ratio of TFFS in selpercatinib versus cabozantinib/vandetanib estimated using a stratified Cox regression model (Cox 1972)

Table JZJB.5. Intercurrent Events of TFFS

Intercurrent Event	IES
Study intervention discontinuation due to reasons other than TFFS event	"Treatment policy" strategy: regardless of whether or not this kind of study intervention discontinuation had occurred
Post study intervention discontinuation anticancer therapy prior to TFFS event	"While on treatment" strategy: consider the assessment of endpoint up until the time that post study intervention discontinuation anticancer therapy is taken
Extended time without adequate assessment prior to TFFS event	"While on treatment" strategy: consider the assessment of endpoint up until the occurrence of extended time without adequate assessment

Abbreviation: IES = intercurrent-event strategies; TFFS = treatment failure-free survival.

Rationale for IES: The interest lies in the treatment effect without the confounding effect of other anticancer therapy or extended time without adequate assessment.

- Study intervention discontinuation due to reasons other than TFFS event is handled with treatment policy as it reflects clinical practice. Time from randomization until disease progression, treatment discontinuation due to unacceptable toxicity or death regardless of other kinds of study intervention discontinuation will be considered in analysis.
- A post study intervention discontinuation anticancer therapy taken prior to TFFS event will confound the treatment effect of selpercatinib in terms of TFFS. If the anticancer therapy is taken, future disease progression/death status is not needed. The participant will be censored and only the time prior to the post study intervention discontinuation anticancer therapy will be considered in analysis.
- TFFS event observed after an extended time without adequate tumor assessment may have occurred much earlier but is not reported because the scheduled assessment was not done. This inadequate observation may introduce bias to TFFS estimates. If extended time without adequate assessment occurs, the participant will be censored and only the time up to the last adequate tumor assessment will be considered in analysis.

8.4.1.1.3. Main Analytical Approach

Since TFFS events include PFS events with additional events of treatment failure due to toxicity, TFFS is sufficiently powered as result of sample size determination based on PFS events.

Conditional on achieving a statistical significance for the primary endpoint of PFS, TFFS by BICR will be tested once at the time of the final analysis which is triggered by the PFS events, in a manner that will preserve the overall Type I error rate at the 1-sided significance level of 0.025 (see Section 8.8 Interim Analyses for details).

- If the study remains at the initial planned total number of PFS events and sample size, the TFFS will be compared between treatment arms using a stratified logrank test, stratified by the 2 randomization strata based on IWRS data: RET mutation (M918T vs. other) and intended control treatment (cabozantinib vs. vandetanib).
- If the total number of PFS events required for the final analysis increases after sample size re-estimation, the TFFS will be compared between treatment arms using the CHW test in a similar manner of the primary efficacy analysis:

Table JZJB.6. Defining the Values for Computing the CHW Test Statistic of TFFS

Quantity	Input Value or Derivation / Calculation
$z_{TFFS,1}$	This is the stratified logrank test statistic comparing TFFS between treatment arms at the interim analysis. The interim analysis is triggered after approximately 56 PFS events have occurred (at 75% PFS information fraction).
$z_{TFFS,2}$	This is the stratified logrank test statistic comparing TFFS between treatment arms at the final analysis.
$e_{TFFS,1}$	This is the observed number of TFFS events across both treatment arms at the interim analysis.
$e_{TFFS,2}$	This is the number of TFFS events across both treatment arms at the final analysis should the study remain at the initial planned 74 total number of PFS events. This number is projected to be 88 assuming a median TFFS of 20 months for Arm B (Wells et al. 2012) and HR of 0.5. $e_{TFFS,2} = 88$
$e^*_{TFFS,2}$	This is the observed number of TFFS events across both treatment arms at the final analysis.

Given the quantities defined in Table JZJB.6, the final CHW test statistic for the TFFS analysis can be written as:

$$z_{TFFS,CHW} = \sqrt{\frac{e_{TFFS,1}}{e_{TFFS,2}}} z_{TFFS,1} + \sqrt{\frac{e_{TFFS,2} - e_{TFFS,1}}{e_{TFFS,2}}} \left\{ \frac{\sqrt{e^*_{TFFS,2}} z_{TFFS,2} - \sqrt{e_{TFFS,1}} z_{TFFS,1}}{\sqrt{e^*_{TFFS,2} - e_{TFFS,1}}} \right\}$$

If $e^*_{TFFS,2} \leq e_{TFFS,2}$, $z_{TFFS,CHW}$ is forced to be equal to $z_{TFFS,2}$, suggesting there is no increase to the initial total number of TFFS events.

The unadjusted HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972). Treatment failure free survival curves, medians, and TFFS rates at various time points with 95% CI for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Table JZJB.7 defines censoring rules to be applied to the TFFS main analysis.

Table JZJB.7. TFFS Censoring Scheme

Situation	Event/Censor	Date of Event or Censor
Tumor progression[a] by BICR, unacceptable toxicity leading to treatment discontinuation per independent review (see Section 8.3.1) or death	Event	Earliest date of PD by BICR, treatment discontinuation or death
No tumor progression by BICR, no unacceptable toxicity leading to treatment discontinuation per independent review, and no death	Censored	Date of last adequate tumor assessment[b], per RECIST 1.1 criteria, or date of randomization (whichever is later)
Unless		
No baseline radiologic tumor assessment available	Censored	Date of randomization
No adequate postbaseline tumor assessment available, no unacceptable toxicity leading to treatment discontinuation per independent review, and death reported after 2 scan intervals[c] following randomization	Censored	Date of randomization
No unacceptable toxicity leading to treatment discontinuation per independent review, and tumor progression by BICR or death documented immediately after 2 or more missing scan intervals following last adequate tumor assessment or randomization (whichever is later)	Censored	Date of last adequate tumor assessment, per RECIST 1.1 criteria, or date of randomization (whichever is later)
New systemic anticancer therapy	Censored	Date of adequate tumor assessment, per RECIST 1.1 criteria, prior to start of new therapy or date of randomization (whichever is later)

Abbreviations: BICR = blinded independent central review; CR = complete response; PD = progressive disease; PR = partial response; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors, version 1.1; SD = stable disease; TFFS = treatment failure-free survival.

[a] Symptomatic deterioration (that is, symptomatic progression that is not radiologically confirmed per RECIST 1.1 criteria) will not be considered as tumor progression.

[b] Adequate tumor assessment per RECIST 1.1 criteria refers to an assessment with 1 of the following responses: CR, PR, SD, or PD.

[c] The 2-scan interval is counted from the date of the last adequate tumor assessment to the date of the next 2 scheduled tumor assessments plus 14 days (adjusted by tumor assessment window).

- If there are multiple dates associated with 1 assessment, the assessment date will be set to the first date when the overall response is PD and the last date otherwise.

8.4.1.1.4. Sensitivity Analyses

Multiple sensitivity analyses for the TFFS will be conducted as defined below:

- counting the time of the third dose reduction due to hypersensitivity, hepatic lab abnormalities, or QT interval corrected for heart rate using Fridericia's formula (QTcF) > 500 msec as the TFFS event for patients on Arm A who continue treatment after > 2 dose reductions (see SA1 in Table JZJB.8);
- ignoring new systemic anticancer therapy (see SA5 Table JZJB.8) according to the EMA guidelines of censoring scheme (EMA 2012);
- discounting the event of PD if pleural effusion is the only reason of PD per BICR and the patient is experiencing chylous effusion at the time of PD (see SA7 Table JZJB.8);

- using other rules for censoring (as defined by Table JZJB.8);
- using an unstratified logrank test and an unstratified Cox regression model;
- using stratification factors based on the eCRF data;
- repeating the main TFFS analysis for the PP population.

Table JZJB.8. TFFS Censoring Scheme - Sensitivity Analysis

Definition	Situation	Event/ Censor	Date of Event or Censor
SA1: Counting the third dose reduction due to hypersensitivity, hepatic lab abnormalities, or QTcF >500 msec as treatment discontinuation	The third dose reduction due to hypersensitivity, hepatic lab abnormalities, or QTcF >500 msec for patients on Arm A who continue treatment after >2 dose reductions	Event	Date of the third dose reduction
SA2: Ignoring absence of adequate postbaseline tumor assessment	No adequate postbaseline tumor assessment available and death reported after 2 scan intervals following randomization	Event	Earliest date of symptomatic deterioration or death
SA3: Ignoring missing tumor assessments	Tumor progression by BICR or death documented after 2 or more missing scan intervals following last adequate tumor assessment or randomization (whichever is later)	Event	Earliest date of PD by BICR or death
SA4: Considering lost to follow up as tumor progression	Patient lost to follow up and no tumor progression by BICR and no death	Event	Date of next scheduled tumor assessment at or after patient was lost to follow up
SA5: Ignoring new systemic anticancer therapy	New systemic anticancer therapy and: 1. No tumor progression by BICR and no death 2. Tumor progression by BICR or death after start of new therapy	1. Censored 2. Event	1. Date of last adequate tumor assessment, per RECIST 1.1 criteria, or date of randomization (whichever is later) 2. Earliest date of PD or death
SA6: Considering events right after the new systemic anticancer therapy starts	Tumor progression by BICR or death within next 14 days from the start of new systemic anticancer therapy	Event	Earliest date of PD by BICR or death
SA7: Discounting PD confounded by chylous effusion	Pleural effusion is the only reason of PD per BICR and the patient is experiencing chylous effusion at the time of PD	Censored	Date of last adequate tumor assessment, per RECIST 1.1 criteria

Abbreviations: BICR = blinded independent central review; PD = progressive disease; QTcF = QT interval corrected for heart rate using Fridericia's formula; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors, version 1.1; SA = sensitivity analysis; TFFS = treatment failure-free survival.

8.4.1.2. Comparative Tolerability

8.4.1.2.1. Research Objective and Question

The research objective for comparative tolerability is specified according to the taxonomy of PRO research objectives by the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) Consortium (Coens et al. 2019) as follows:

- to demonstrate superior tolerability of Arm A as compared to Arm B in patients with progressive, advanced, kinase inhibitor naive, RET-mutant MTC.

The associated research question is:

- what is the difference in the mean proportion of time on treatment with high side effect bother among patients treated in Arm A compared to those treated in Arm B?

8.4.1.2.2. Definition of Endpoint

The estimand for this secondary objective is described by the following attributes:

- Population: patients with progressive, advanced, kinase inhibitor-naïve, RET-mutant MTC. Further details can be found in Section 8.4.1.2.3.
- Endpoint:
  - A single item "I am bothered by side effects of treatment" from the Functional Assessment of Cancer Therapy (FACT) general scale will be utilized to assess overall treatment side-effect burden. The Functional Assessment of Cancer Therapy-Side Effects (FACT-GP5) item is measured on a 5-item Likert scale: 0 (not at all); 1 (a little bit); 2 (somewhat); 3 (quite a bit); or 4 (very much). This item will be completed weekly by the patient via an electronic clinical outcomes assessment (eCOA) device provided; sites will not administer this instrument.
  - Comparative tolerability is defined as a comparison of the proportion of time on treatment with a high symptom burden as assessed by the FACT-GP5 between Arm A and Arm B. High symptom burden is defined as a FACT-GP5 score of 3 or 4 (ie "Quite a bit" or "Very much" on a 5-point Likert scale). This proportion will be calculated for each patient using:
    - As numerator - the cumulative amount of time, in weeks, during which a patient reports high side effect burden (ie selects a score of 3 or 4 on the FACT-GP5 response scale).
    - As denominator - the total duration of therapy (weeks), derived as (date of last study treatment dose - date of first study treatment dose + 1) ÷ 7.
- Treatment condition: the randomized study intervention (selpercatinib or cabozantinib/vandetanib) will be administered twice or once a day in continuous 28-day cycles until disease progression, unacceptable toxicity or other protocol-defined reasons for treatment discontinuation. Further details on study interventions including interventions, concomitant therapy and dose modification can be found in Protocol Section 6 Study Intervention.
- Intercurrent-event strategies for main analytical approach are listed in Table JZJB.9.

- Population-level summary measure: difference in the proportion of time with high side effect bother in selpercatinib versus cabozantinib/vandetanib

Table JZJB.9. Intercurrent Events of Comparative Tolerability

Intercurrent Event	IES
Treatment discontinuation for any reason (including disease progression or death)	"While on treatment" strategy: the proportion of time with high side effect bother is calculated using treatment exposure in the denominator of the endpoint, so any treatment discontinuation is taken into account by design
Nonadherence to the study treatment	"Treatment policy" strategy: no modification will be made to the calculation of the proportion of time with high side effect bother to take into account the adherence of the patient to the treatment

Abbreviation: IES = Intercurrent-Event Strategies.

8.4.1.2.3. Main Analytical Approach

Comparative tolerability is also a key secondary endpoint and will be tested conditionally on achieving a statistical significance for PFS and TFFS by BICR to preserve the overall Type I error rate at the 1-sided significance level of 0.025 (see Section 8.8 Interim Analyses for details).

The proportion of time with high side effect bother will be compared between treatment arms using the van Elteren test (van Elteren 1960), a stratified Wilcoxon rank sum test (Wilcoxon 1946), stratified by the 2 randomization strata based on IWRS data: RET mutation (M918T vs. other) and intended control treatment (cabozantinib vs. vandetanib).

Comparative tolerability will be analyzed in the Tolerability Evaluable Population as defined in Section 7, ie patients enrolled by the time of the interim analysis (approximately 250 patients projected to be enrolled) regardless of the sample size re-estimation result. Since the sample size is fixed for the comparative tolerability endpoint, no adjustment of test statistic will be applied and the Type I error will not be inflated by a potential sample size increase.

8.4.1.2.4. Missing Data Handling

Missing data can occur at 2 levels in this analysis:

- at the individual FACT-GP5 assessment time points, and
- at the patient level if there are too many consecutive, missing FACT-GP5 assessments for a given patient (making the calculation of the proportion of time with high side effect bother for this patient too uncertain).

Missing individual FACT-GP5 assessments will be handled in the calculation of the proportion of time with high side effect. The following rules will be applied to calculate the proportion in case of missing FACT-GP5 assessments:

1. If 1 or 2 consecutive FACT-GP5 assessments are missing, they will be imputed according to the available FACT-GP5 assessments as follows:
    - If both the FACT-GP5 assessments immediately before and immediately after the missing assessment(s) indicate no high side effect bother, then the missing assessment(s) will be considered as indicating no high side effect bother.

- If any 1 of the FACT-GP5 assessments immediately before or immediately after the missing assessment(s) indicate high side effect bother, then the missing assessments will be considered as indicating high side effect bother.
- A missing FACT-GP5 at baseline will not be imputed nor will it inform the imputation of other missing values, as it relates to the period before treatment start and is not included in the calculation of proportion of time with high side effect bother.
- For patients for whom Cycle 1 Day 8 is missing, only Cycle 1 Day 15 will be used to impute the missing value. If both Cycle 1 Day 8 and Cycle 1 Day 15 are missing, only Day 15 can be imputed from Day 21. Day 8 will not be imputed and will be considered missing.

2. If 3 or more consecutive FACT-GP5 assessments are missing, the period corresponding to these missing assessments will be subtracted from both the numerator and denominator
3. If FACT-GP5 assessments are missing before treatment discontinuation, they will be imputed as follows:

    - If 1 or 2 FACT-GP5 assessments are missing immediately before discontinuation, they will be given the same value as indicated by the last available FACT-GP5 assessment.
    - If 3 or more FACT-GP5 assessments are missing immediately before discontinuation, then the period corresponding to these missing assessments will be subtracted from both the numerator and denominator.

The distribution of missing data will be summarized by treatment arm. The impact of missing data on the results at patient level will be explored in the sensitivity analyses.

8.4.1.2.5. Sensitivity Analyses

8.4.1.2.5.1. Impact of Threshold Used to Define High Side Effect Bother

In the main analysis of the comparative tolerability endpoint, the proportion of time with high side effect bother is calculated under the assumption that a FACT-GP5 score of ≥3 characterizes high side effect bother for the patient. Sensitivity analyses will be conducted to explore the impact of this assumption on the conclusion of the comparison between the treatment arms by considering different thresholds for the FACT-GP5 score for the calculation. The analyses will be replicated with the following calculations of proportion of time: proportion of time with a FACT-GP5 score ≥2 and proportion of time with a FACT-GP5 score of 4.

8.4.1.2.5.2. Impact of Imputation Rule for Missing FACT-GP5 Assessment in Calculation of Proportion of Time with High Side Effect Bother

The calculation of proportion of time with high side effect bother, as specified above for the main analyses, includes a rule for imputing missing FACT-GP5 assessments. The impact of this imputation rule will be explored by replicating the analysis using alternative approaches for the management of missing FACT-GP5 assessments in the calculation of the proportion of time with high side effect bother. The following imputation strategy will be applied:

- **Most conservative imputation strategy** - any missing FACT-GP5 assessment is considered indicative of high side effect bother.
- **Least conservative imputation strategy** - any missing FACT-GP5 assessment is considered indicative of no high side effect bother.
- **No imputation** - No imputation is made for any missing FACT-GP5 assessments; all periods of missing data are excluded from both the numerator and denominator.

8.4.1.2.5.3. Impact of Patients with Prolonged Periods of Consecutive Missing FACT-GP5 Assessments

In the main analysis of the comparative tolerability endpoint, all patients of the Tolerability Evaluable Population will be included. However, for some patients, due to prolonged periods of consecutive missing FACT-GP5 assessments, side effect bother cannot be imputed per the rules defined in Section 8.4.1.2.3 for every missing FACT-GP5 assessments over the course of treatment. If patients have very high missingness, inclusion of the patients may impact the interpretation of outcomes. To evaluate the impact of patients with prolonged periods of consecutive missing FACT-GP5 assessments, sensitivity analyses will be performed by replicating the main analysis in the following subpopulations of the Tolerability Evaluable Population patients:

- for whom side effect bother, as assessed by the FACT-GP5 item, is available or can be imputed for ≥50% of the total duration of treatment.
- for whom side effect bother, as assessed by the FACT-GP5 item, is available or can be imputed for ≥80% of the total duration of treatment.
- for whom side effect bother, as assessed by the FACT-GP5 item, is available or can be imputed for ≥90% of the total duration of treatment.
- for whom side effect bother, as assessed by the FACT-GP5 item, is available or can be imputed for 100% of the total duration of treatment.

8.4.1.2.6. Supplementary Analyses

8.4.1.2.6.1. Comparison of Selpercatinib Versus Cabozantinib and Vandetanib Separately

A subgroup analysis will be conducted based on a stratification factor of intended treatment if subsequently randomized to control arm to estimate the difference in proportion of time with high side effect bother between selpercatinib and each of the 2 treatment options of the control arm. The same analytical approach will be used as for the main analysis.

8.4.1.2.6.2. Other Analytical Approaches

No prior FACT-GP5 data of MTC patients is available to inform the distribution of proportion of time with high side effect bother. Without a valid assumption of the distribution, nonparametric approaches are used in the main analysis. Depending on the empirical distribution observed at the time of analysis, multivariable linear regressions with original value and with proper transformations of proportion of time with high side effect bother (eg logit transformation) will be performed as supplementary analyses. The difference in mean proportion of time with high

side effect bother between treatment arms with 95% CI will be estimated using the multivariable linear regressions. Proportion of time with high side effect bother (with or without transformation) will be used as the outcome variable, treatment group (Arm A vs. Arm B) as the explanatory variable, and the 2 randomization strata as covariates. Other advanced analytical approaches may also be performed as deemed appropriate.

8.4.2. Supportive Secondary Endpoints

Treatment Failure-free Survival per Investigator Assessment

Treatment failure-free survival per investigator assessment is defined as the time from randomization to the first occurrence of documented radiographic disease progression per RECIST 1.1 as assessed by the investigator or unacceptable toxicity leading to treatment discontinuation as assessed by the investigator (regardless if a study drug-related AE meets protocol guidance for treatment discontinuation or not); or death (due to any cause). Treatment failure-free survival per investigator assessment will be analyzed using the same methodology as for the TFFS per BICR.

Progression-free Survival per Investigator Assessment

Progression-free survival per investigator assessment is defined according to the same criteria and will be analyzed using the same methodology as for the PFS per BICR.

One more censoring rule will be implemented as a part of sensitivity analyses: considering symptomatic deterioration as tumor progression. In this situation the earliest date of PD per RECIST 1.1 criteria, symptomatic deterioration or death will be assigned as an event date.

Overall Response Rate and Duration of Response

Overall response rate (ORR) is defined as the number of patients who achieve a best overall response (BOR) of CR or partial response (PR) divided by the total number of patients randomized to each treatment arm. Best overall response is determined from a sequence of responses assessed. Two objective status determinations of CR before progression are required for a best response of CR. Two determinations of PR or better before progression, but not qualifying for a CR, are required for a best response of PR. The second objective response is required to be ≥28 days after the initial response. Tumor assessments performed after initiation of new anticancer treatment (systemic therapy) will be excluded from evaluating the BOR. Patients who do not have any post baseline tumor response assessments are considered nonresponders and are included in the denominator when calculating the response rate.

Disease control rate (DCR) is defined as the number of patients who achieve a BOR of CR, PR or stable disease (SD) lasting 16 or more weeks divided by the total number of patients randomized to each treatment arm.

The ORR and DCR, with 95% CI, will be summarized for each treatment arm. Overall response rate and DCR will be compared between Arm A and Arm B using a Cochran-Mantel-Haenszel (CMH) test stratified by the randomization strata. The ORR and DCR according to both BICR and investigator-assessed BOR will be evaluated.

Maximum change in tumor size is defined as the ratio of best postbaseline tumor size over that of baseline. The maximum reduction from baseline in the sum of target lesions (based on investigator assessment) will be presented per patient in a waterfall plot.

Duration of response (DOR) is defined as the time from the date that measurement criteria for CR or PR (whichever is first recorded) are first met until the first date that disease is recurrent or documented disease progression is observed, per RECIST 1.1 criteria, or the date of death from any cause in the absence of documented disease progression or recurrence. The DOR according to both BICR and investigator-assessed BOR will be evaluated. The DOR will be analyzed for patients who achieve a BOR of CR or PR.

Duration of response will be compared between treatment arms using a stratified logrank test, stratified by the randomization strata. The corresponding HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972), stratified by the randomization strata. Survival curves, the median and rates at various time points with 95% CI, for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Overall Survival

Overall survival (OS) is defined as the time from randomization until death from any cause. If the patient is alive or lost to follow-up at the time of data analysis, OS data will be censored on the last date the patient is known to be alive.

The study is not powered for OS but will continue to approximately 125 OS events or until a maximum of 6 years from the first patient visit, whichever comes first. If the true OS HR is 0.7, 125 OS events will provide approximately 98% probability that the observed OS HR would be less than 1.0, indicating no OS detriment associated with the selpercatinib treatment. A HR of 0.7 translates into a 42.9% relative and 30 months absolute increase in median OS (if assuming a median OS of 70 months for cabozantinib/vandetanib and 100 months for selpercatinib).

Overall survival will be compared between treatment arms using a stratified logrank test, stratified by the randomization strata. The corresponding HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972), stratified by the randomization strata. Overall survival curves, the median and OS rates at various time points with 95% CI, for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Multiple sensitivity analyses for the OS will be conducted as defined below:

- using an unstratified logrank test and an unstratified Cox regression model
- using stratification factors based on the eCRF data
- repeating the main OS analysis for the PP population
- censoring OS at start of new systemic anticancer therapy including crossover therapy for patients assigned to Arm B who crossed over to selpercatinib
- censoring OS for crossover patients at start of crossover therapy for patients assigned to Arm B who crossed over to selpercatinib
- censoring OS for patients who continued study therapy beyond progression at the earliest date of study therapy after tumor progression by BICR.

Progression-free Survival 2

Progression-free survival 2 is defined as the time from randomization to disease progression on the next line of treatment or death from any cause in the absence of observed disease progression. If the patient is alive at the cutoff date for the analysis, and disease progression has not been observed, PFS2 data will be censored on the latest date of last progression-free assessment or start of the next line of treatment. Table JZJB.10 defines censoring rules to be applied to the PFS2 analysis.

Progression-free survival 2 will be compared between treatment arms using a stratified logrank test, stratified by the randomization strata. The corresponding HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972), stratified by the randomization strata. Progression-free survival 2 curves, the median and PFS2 rates at various time points with 95% CI, for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Table JZJB.10. PFS2 Censoring Scheme

Situation	Event/Censor	Date of Event or Censor
Tumor progression[a] or death on the next line of treatment	Event	Earliest date of PD or death
No tumor progression and no death on the next line of treatment	Censored	Latest date of last progression-free assessment or start of the next line of treatment
	Unless	
No baseline radiologic tumor assessment available	Censored	Date of randomization
Patient is still on study treatment	Censored	Date of last progression-free assessment
Patient discontinued the study treatment but has not started next line of treatment or Patient had tumor progression on study treatment but has not discontinued		
1. death 2. patient is alive	1. Event 2. Censored	1. Date of death 2. Last date the patient is known to be alive
No tumor progression prior to the start of the next line of treatment	Censored	Date of last progression-free assessment prior to the start of the next line of treatment

Abbreviations: PD = progressive disease; PFS = progression-free survival; PFS2 = progression-free survival 2; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors, version 1.1.

[a] Symptomatic deterioration (that is, symptomatic progression that is not radiologically confirmed per RECIST 1.1 criteria) will be considered as tumor progression.

- If there are multiple dates associated with 1 assessment, the assessment date will be set to the first date when the overall response is PD and the last date otherwise.

Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) derived from the verbatim term will be used when reporting AEs by MedDRA terms. The MedDRA Lower Level Term (LLT) will be used in the treatment-emergent computation. Severity will be measured using the grade defined by the National Cancer Institute - Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 5.0.

Treatment-emergent adverse events (TEAEs) are events that first occurred or worsened in severity after baseline and up to 30 days after core period treatment discontinuation.. Treatment-emergent adverse events will be summarized by System Organ Class (SOC) and by decreasing frequency of PT within SOC. TEAE after crossover will be defined and summarized separately in Section 8.5.

Serious adverse events (SAEs) are any AEs that result in any of the following outcomes:

- death
- initial or prolonged inpatient hospitalization
- a life-threatening experience (ie, immediate risk of dying)
- persistent or significant disability/incapacity
- congenital anomaly/birth defect
- important medical events (detailed description is contained in the protocol)

Adverse event analyses will include summaries of the following:

- AEs, including severity and possible relationship to study drug
- SAEs, including possible relationship to study drug
- AEs leading to dose adjustments
- discontinuations from study treatment due to AEs or death
- time to any toxicity grade 3 or above
- **AEs of special interest** (AESI): Categories of AESI will be identified as the understanding of the safety of selpercatinib increases. The final list of categories will be maintained at both compound and study level and reported in the Clinical Study Report.
- **Consolidated AEs** are composite AE terms consisting of synonymous PTs to allow meaningful interpretation of the AE data. The final list of consolidated AE categories and PTs will be maintained at both compound and study level and reported in the Clinical Study Report.

Deaths

- Summary of deaths (all deaths and deaths within 30 days of last dose of study drug) and their primary cause (study disease progression, AE, other)
- Listing of TEAEs leading to death.

Laboratory Abnormalities

The severity of laboratory results will be classified according to NCI-CTCAE. Treatment-emergent changes in laboratory values will be reported for the safety population. The laboratory toxicity by worst NCI-CTCAE grade and shifts in toxicity grading from baseline to the worst postbaseline grade will be summarized.

Shift to low/high tables will include the number and percentage of patients within each baseline category (baseline value is low, normal, high, or missing) versus each postbaseline category (worst value is low, normal, high, or missing) by treatment arm.

RET Testing Results Concordance

The concordance between RET testing results based on local laboratory tests and centrally assessed results will be evaluated. Details regarding analyses involving centrally assessed RET testing results will be described in a separate diagnostics SAP.

8.5. Tertiary/Exploratory Endpoint Analyses

Calcitonin response rate

Calcitonin response rate is defined as percentage of patients who had a decline from baseline in the calcitonin level of at least 50% maintained for at least 4 weeks. Calcitonin response rate will be compared between Arm A and Arm B using a CMH test stratified by the randomization strata.

CEA response rate

Carcinoembryonic antigen (CEA) response rate is defined as percentage of patients who had a decline from baseline in the CEA level of at least 50% maintained for at least 4 weeks. Carcinoembryonic antigen response rate will be compared between Arm A and Arm B using a CMH test stratified by the randomization strata.

PROs

Further details of planned analysis of PROs will be provided in a separate SAP.

PFS/ORR/DOR/DCR After Crossover

Progression-free survival after crossover in crossover population is defined as the time from start of selpercatinib treatment until the occurrence of disease progression or death from any cause. Patients known to be alive and without disease progression will be censored at the time of the last adequate tumor assessment.

Duration of response is defined as the time from the date that measurement criteria for CR or PR with respect to the crossover baseline tumor assessment are first met until the first date that disease is recurrent or documented disease progression is observed, or the date of death from any cause in the absence of documented disease progression or recurrence during crossover period.

Survival curve, the median, and rates at various time points with 95% CI will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Overall response rate after crossover is defined as the number of patients who achieve a BOR of CR or PR with respect to the crossover baseline tumor assessment, divided by the total number of patients in crossover population. Disease control rate after crossover is defined as the number of patients who achieve a BOR of CR, PR or stable disease (SD) lasting 16 or more weeks, with respect to the crossover baseline tumor assessment, divided by the total number of patients in crossover population. The ORR and DCR after crossover with 95% CI will be summarized.

Additional exploratory analyses to compare efficacy endpoints prior versus post crossover intra-patient may be performed as deemed appropriate.

TEAE and Lab Abnormalities After Crossover

Treatment-emergent adverse events (TEAEs) after crossover are events that first occurred or worsened in severity after crossover baseline and up to 30 days after crossover treatment discontinuation. Treatment-emergent adverse events will be summarized by System Organ Class (SOC) and by decreasing frequency of PT within SOC.

The severity of laboratory results after crossover will be classified according to NCI-CTCAE. Treatment-emergent changes in laboratory values after crossover will be reported for the crossover population. The laboratory toxicity by worst NCI-CTCAE grade and shifts in toxicity grading from crossover baseline to the worst post crossover baseline grade will be summarized.

Time to Initiation of New Anticancer Therapy

Time to initiation of new anticancer therapy is defined as the time from randomization to documentation in the case report forms of a new local (eg surgery, radiation) or systemic anticancer treatment administered to the patient.

Death is considered as a competing event. To account for the competing risk from death, the probability of initiation of new anticancer therapy will be estimated by treatment arm using a cumulative-incidence function (subdistribution function). A plot of the estimated cumulative-incidence functions will be displayed for each treatment arm. Time to initiation of new anticancer therapy will be compared between treatment arms using a stratified logrank test, computed on the basis of cause-specific hazard functions. The corresponding cause-specific HRs will be estimated using a stratified Cox regression model.

Biomarkers

Biomarkers assessed from blood or tissue samples and their relationship with clinical outcomes will be analyzed according to a separate translational research analysis plan.

8.6. (Other) Safety Analyses

8.6.1. Extent of Exposure

The number of cycles received, dose reductions, dose withholding, and dose intensity for each drug will be summarized for all treated patients by treatment arm. The exposure derivations are provided below:

- Duration of therapy (weeks) = (date of last dose of selpercatinib - date of first dose of selpercatinib + 1) ÷ 7
- Cumulative dose (mg) = sum of all doses
- Dose intensity (mg/week) = (cumulative dose) ÷ (duration of therapy)
- Planned dose intensity (mg/week) = daily dose (mg) * 7
- Relative dose intensity (%) = (dose intensity / planned dose intensity) * 100

All analyses of selpercatinib exposure will be performed for the study treatment period for patients randomly assigned to Arm A as well as for the crossover period for patients assigned to Arm B who crossover to selpercatinib after disease progression. All analyses of cabozantinib and vandetanib exposure will be performed for the study treatment period for patients randomly assigned to Arm B and stratified to receive cabozantinib or vandetanib respectively.

8.6.1.1. Treatment Compliance

Compliance for each drug will be assessed as the proportion of treatment that is actually taken, relative to what is expected, after accounting for protocol-defined dose adjustments. Study treatment taken will be derived from the difference between the total number of capsules dispensed and returned over the course of the patient's treatment. A patient will be considered noncompliant if he or she takes <75% or ≥125% of the planned doses for the duration of the study treatment.

Summaries will be provided for the study treatment period for Arm A and Arm B patients as well as for the crossover period for patients assigned to Arm B who crossover to selpercatinib after disease progression.

8.6.2. Additional Safety Assessments

Electrocardiograms

Summary will be provided for patient who experienced QTcF >500 msec on at least 2 of 3 electrocardiograms (ECG) and had the triplicate average QTcF greater than 500 msec.

Abnormal changes (also known as delta changes) of QTcF >60 msec will also be summarized. The changes will be derived as from maximum baseline QTcF interval to maximum QTcF interval during treatment.

Vital Signs

Treatment-emergent abnormal changes in vital signs will be summarized using descriptive statistics.

8.7. Other Analyses

8.7.1. Demographic and Other Baseline Characteristics

Demographic and baseline patient and disease characteristics, historical diagnoses, preexisting conditions, and prior therapies will be reported for the ITT population using descriptive statistics.

8.7.2. Concomitant Therapy

A summary of preferred names of concomitant medication by treatment arm by decreasing frequency will be generated.

8.7.3. Poststudy Treatment Therapy

The numbers and percentages of patients receiving poststudy anticancer therapies will be provided by type of therapy (surgery, radiotherapy, or systemic therapy), by drug class and/or name, overall, and by line of therapy.

8.7.4. Follow-up Time

Follow-up time is defined as the time from the date of randomization until death from any cause or last date the patient is known to be alive and under follow-up. Follow-up time will be estimated using Kaplan-Meier estimation of potential follow-up ("reverse Kaplan-Meier") (Schemper and Smith 1996). The inverse of the censoring rules for the OS will be used, ie considering all censoring times for OS as event times (times when the patient is known to be still alive and under follow-up) and censoring patients at the date of death.

Follow-up time will be compared across treatment arms using an unstratified logrank test and Kaplan-Meier estimates (Kaplan and Meier 1958).

8.7.5. Health Care Resource Utilization

Summaries of hospitalizations, emergency room visits, granulocyte-colony stimulating factor (G-CSF) use, transfusions, and analgesic use will be reported descriptively for each treatment arm. Due to potential differential treatment durations, hospitalizations and emergency room visits per on-treatment month will also be reported by arm.

Duration of hospital stays and average number of emergency room visits will be reported by treatment arm.

8.7.6. Pharmacokinetic/Pharmacodynamic Analyses

Selpercatinib plasma concentrations will be summarized by descriptive statistics. Additional analysis utilizing the population pharmacokinetics (PK) approach may also be conducted if deemed appropriate and described in a separate SAP. The relationship between selpercatinib plasma exposure and selected efficacy and safety outcomes may be explored.

8.7.7. Subgroup Analyses

To determine whether the treatment effect is consistent across various subgroups, the between-group treatment effect for OS, PFS, TFFS and ORR (with a nominal 95% CI) will be estimated and plotted within each category of the following subgroups (defined based on eCRF data):

- Age category (≤65 vs. >65 years)
- Eastern Cooperative Oncology Group (ECOG) performance status (0 to 1 vs. 2)
- Sex (female vs. male)
- Race (Asian vs. non-Asian)
- Tissue vs. blood RET mutation detection
- RET mutation: M918T vs. other
- RET mutation: M918T vs. extracellular cysteine region (mutations in the exons 10 and 11) vs. other
- Investigator's choice of treatment with cabozantinib vs. vandetanib

If a level of a factor includes fewer than 5% of the ITT population, analysis within that level will be omitted. Additional subgroup analyses may be performed as deemed appropriate.

8.7.8. Analyses of Pandemic Mitigations

There may be times due to exceptional circumstances (eg, coronavirus disease 2019 [COVID-19] pandemic) where it may not be feasible for patients to come to investigator sites for study-required visits. To evaluate the impact of pandemic mitigations, missing visit/values will be summarized. Additional sensitivity analyses may be performed as deemed appropriate.

8.7.9. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements. Analyses provided for the CTR requirements include the following:

- Summary of AEs, provided as a dataset which will be converted to an extensible markup language (XML) file. Both SAEs and "Other" AEs are summarized by treatment arms and by MedDRA PT.
- An AE is considered "Serious" whether or not it is a TEAE.
- An AE is considered in the "Other" category if it is both a TEAE and is not serious.
- For each SAE and "Other" AE, for each term and treatment arms, the following are provided:
    - the number of participants at risk of an event (if certain subjects cannot be at risk for some reason, for example, gender-specific AEs, then the number will be adjusted to only include the patients at risk)
    - the number of participants who experienced each event term
    - the number of events experienced.
- For each SAE, for each term and treatment arms, the following are also provided for the EudraCT results submission:
    - The number of occurrences (events) causally related to treatment
    - The total number of deaths
    - The number of deaths causally related to treatment

- Consistent with www.ClinicalTrials.gov requirements, a threshold for frequency of "Other" AEs can be implemented rather than presenting all "Other" AEs. For example, "Other" AEs that occur in fewer than 5% of patients in any treatment arms may not be included if a 5% threshold is chosen. The frequency threshold must be less than or equal to the allowed maximum of 5%.
- A participant flow will be created that will describe:
    - Number of participants per treatment arm. Screen failures do not need to be included. Number of participants who did not complete the study per treatment arm. This analysis will be based on study discontinuation, not treatment discontinuation.
    - Reasons participants did not complete the study.

8.8. Interim Analyses

An interim analysis will be triggered after approximately 56 PFS events by BICR have occurred (at 75% information fraction). Based on the observed HR of PFS by BICR and pre-specified criteria, the study results at interim analysis will be categorized into 4 determination zones: efficacious, favorable, promising, and unfavorable.

- If the study results fall into the **efficacious** zone, a positive study will be declared due to overwhelming efficacy, enrollment will be stopped, and regulatory interactions will be initiated.
- If the study results fall into the **favorable** zone, the study will continue to the final analysis with the initially planned total number of events at 74 (ie, no re-estimation will be conducted).
- If the study falls into the **promising** zone, the total number of events will be increased. The events re-estimation will be based on the observed PFS HR according to the method provided in Cui, Hung, and Wang (Cui et al. 1999). The total number of patients will be increased and determined accordingly so that the required number of events could be achieved in a desired timeframe.
- If the study results fall into the **unfavorable** zone, enrollment will be halted, and the study will continue to the final analysis.

The critical boundary of efficacious zone to declare a positive study is determined by the Rho family alpha spending function with Rho=12 (Kim and DeMets 1987), based on consideration of being more conservative than O'Brien Fleming alpha spending function in stopping at the interim analysis. The one-sided alpha spent at the interim analysis is 0.00088. The p-value, Z scale, and HR critical boundaries are shown in Table JZJB.11 for reference only and will be updated based on the actual number of events observed at the time of interim and final analysis.

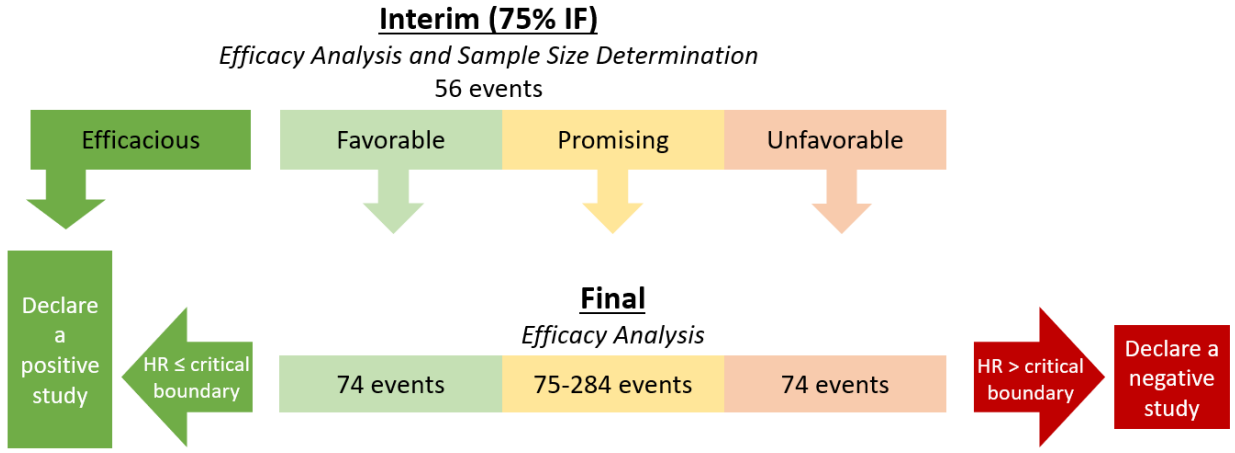
The decision boundary of other zones and the algorithm of sample size re-estimation will be described in the Adaptive Design Charter.

Table JZJB.11. Efficacy Analyses

Efficacy Analysis	Information Fraction	Number of PFS Events by BICR to Trigger Analyses	1-sided P-value Critical Boundary	Z Scale Critical Boundary	HR Critical Boundary
Interim	75%	56	0.00088	-3.127	0.412
Final	100%	74 – 284	0.02497	-1.960	TBD

Abbreviations: BICR = Blinded Independent Central Review; HR = Hazard Ratio; PFS = Progression-Free Survival; TBD = To Be Determined.

The figure below provides an overview of the proposed adaptive design. Sample size/total number of events determination will be conducted at the interim efficacy analysis. If the study is not declared positive, final analysis will be conducted according to a determination zone (favorable, promising, or unfavorable) where the interim results fall. At final analysis, the study will be declared positive if the critical boundary is met.



Abbreviations: IF=Information Fraction; HR=Hazard Ratio

Figure JZJB.1 Overview of Proposed Adaptive Design

Conditional on PFS by BICR achieving statistical significance, TFFS by BICR will be tested and then comparative tolerability. Both TFFS and comparative tolerability will be formally tested at most once at final analysis. A nominal alpha of 0.00001 will be spent for interim analysis and both key secondary endpoints will be tested against 1-sided significance level of 0.02499 at final analysis.

8.8.1. Maintaining the Trial Integrity

8.8.1.1. Data Monitoring Committee

This is an open-label study where participants and investigators will not be blinded. However, to preserve the integrity of the trial Lilly will not have unblinded access to aggregate data from the clinical database until the database lock for the final analysis. An independent Data Monitoring Committee (iDMC) will monitor safety-related data during the course of the trial on a regular basis. In order to minimize the operational and statistical bias that result from performing an interim efficacy analysis with sample size re-estimation, it will be conducted under the auspices of an iDMC. The purpose of the iDMC is to advise Lilly regarding the required total number of events, the continuing safety of study participants, and the continuing validity and scientific merit of the trial.

An external Statistical Analysis Center (SAC) will conduct interim efficacy and safety analyses and report the results to the iDMC. This practice will ensure that Lilly personnel involved in the day-to-day management and conduct of the trial do not have access to unblinded comparative results, even inadvertently.

Details are contained in a separate iDMC charter.

8.8.1.2. Actions by Sponsor

- To avoid disclosure of interim efficacy results conveyed by adaptation decisions, the pre-specified adaptation algorithm and details on the sample size/events number determination will be preserved in an Adaptive Design Charter and stored confidentially in a restricted access area. Only the iDMC, SAC, and the Lilly statisticians involved in the adaptive design will have access to the Adaptive Design Charter. No other study team members will have access.
- Communication plan will be developed and documented prior to the interim analysis, including:
    - Any members of the study team that have access to the interim analysis / adaptation decision
    - Roles and responsibilities of who will communicate decisions/outcomes to whom
- No information about trial screening, enrollment, and patient progression will be shared with the sites. In addition, the ERB, investigators, and trial participants will be shielded from knowledge of adaptive changes. For example, if the sample size / total number of events are increased after the interim efficacy analysis, they will be informed that the targeted patient/event number hasn't been reached rather than being notified of the specific targeted final sample size / number of events. Lilly personnel who have direct interaction with study sites will be appropriately trained on this expectation.
- Study level unblinding plan is developed and describes who has access to patient level unblinded data, who has access to blinded or unblinded aggregate data. Details are contained in a separate blinding and unblinding plan. Documentation of who accessed blinding or unblinded aggregate data will be maintained.
- Aggregate data will be transferred to CLUWE containers with restricted access set up as described in the blinding and unblinding plan.

8.8.2. Safety Update Report

The following reports may be needed for the Development Safety Update Report (DSUR), Periodic Safety Update Report (PSUR), Japan Periodic Safety Review (Japan-PSR), etc.:

- Estimated cumulative subject exposure
- Cumulative exposure to investigational drug by demographic characteristics
- Listing of subjects who died during the report period
- Listing of discontinuations due to AEs during the report period including both core study and crossover periods
- Summary of overall treatment-emergent adverse events (TEAE) including both core study and crossover periods
    - Overall TEAE are events that first occurred or worsened in severity after baseline and up to 30 days after the latest study treatment discontinuation, including either core study or crossover period.

9. References

Coens C, Madeline P, Basch E, et al. Improving standards of patient reported outcomes analysis: matching statistical methods to key research objectives-a SISAQOL initiative. *Qual Life Res*. 2019;28:S183-S184.

Cox DR. Regression models and life-tables. *J Royal Stat Soc Ser B*. 1972;34(2):187-220.

Cui L, Hung HMJ, Wang SJ. Modification of Sample Size in Group Sequential Clinical Trials. *Biometrics*. 1999;55(3):853-857. https://doi.org/10.1111/j.0006-341X.1999.00853.x

Elisei R, Schlumberger MJ, Müller SP, et al. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol*. 2013;31(29):3639-3646. https://doi.org/10.1200.JCO.2012.48.4659

Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45(2):228-247. https://doi.org/10.1016/j.ejca.2008.10.026

[EMA] European Medicines Agency. Appendix 1 to the guideline of the evaluation of anticancer medicinal products in man. EMA/CHMP/27994/2008/Rev.1. Published December 13 2012. Accessed October 2021. https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481. http://doi.org/10.1080/01621459.1958.10501452

Kim K, DeMets DL. Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika*. 1987;74(1): 149-154. https://doi.org/10.1093/biomet/74.1.149

Kreissl MC, Bastholt L, Elisei R, et al. Efficacy and Safety of Vandetanib in Progressive and Symptomatic Medullary Thyroid Cancer: Post Hoc Analysis From the ZETA Trial. *J Clin Oncol.* 2020;38(24):2773-2781. https://doi.org/10.1200/jco.19.02790

Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials*. 1996;17(4):343-346. https://doi.org/10.1016/0197-2456(96)00075-x

Schlumberger M, Elisei R, Müller S, et al. Overall survival analysis of EXAM, a phase III trial of cabozantinib in patients with radiographically progressive medullary thyroid carcinoma. *Ann Oncol*. 2017;28(11):2813-2819. https://doi.org/10.1093/annonc/mdx479

Sherman SI, Clary DO, Elisei R, et al. Correlative analyses of RET and RAS mutations in a phase 3 trial of cabozantinib in patients with progressive, metastatic medullary thyroid cancer. *Cancer*. 2016;122(24):3856-3864. https://doi.org/10.1002/cncr.30252

van Elteren P. On the combination of independent two sample tests of Wilcoxon. *Bulletin of the Institute of International Statistics*. 1960;37(3):351–361.

Wilcoxon F. Individual comparisons of grouped data by ranking methods. *J Econ Entomol*. 1946;39:269. https://doi.org/10.1093/jee/39.2.269

Wells SA Jr, Robinson BG, Gagel RF, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol*. 2012;30(2):134-141. https://doi.org/10.1200/JCO.2011.35.5040

10. Appendices

Appendix 1. List of Abbreviations

Term	Definition
AE	Adverse event
AESI	Adverse event of special interest
BICR	Blinded independent central review
BOR	Best overall response
CEA	Carcinoembryonic antigen
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CONSORT	Consolidated Standards of Reporting Trials
COVID-19	Coronavirus disease 2019
CR	Complete response
CTCAE	Common Terminology of Criteria in Adverse Events
CTR	Clinical Trial Registry
iDMC	Independent Data Monitoring Committee
DNA	Deoxyribonucleic acid
DOR	Duration of response
DSUR	Development Safety Update Report
ECOG	Eastern Cooperative Oncology Group
ECG	Electrocardiograms
eCOA	Electronic clinical outcomes assessment
ECOG	Eastern Cooperative Oncology Group
EORTC QLQ-30-C30-PF	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 Version 3.0
EORTC IL19	European Organisation for Research and Treatment of Cancer, item library 19
eCRF	electronic case report form

Term	Definition
FACT	Functional Assessment of Cancer Therapy
FACT-GP5	Functional Assessment of Cancer Therapy-Side Effects
G-CSF	granulocyte-colony stimulating factor
HR	hazard ratio
HRQoL	health-related quality of life
IPD	important protocol deviation
ITT	intention to treat
IWRS	interactive web-response system
LLT	Lower Level Term
MedDRA	Medical Dictionary for Regulatory Activities
MTC	medullary thyroid cancer
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
ORR	overall response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PFS2	progression-free survival 2
PK	pharmacokinetic
PP	per-protocol
PR	partial response
PRO	patient-reported outcome
PT	preferred term
QTcF	QT interval corrected for heart rate using Fridericia's formula
RECIST	Response Evaluation Criteria in Solid Tumors
RET	rearranged during transfection
SA	sensitivity analysis

Term	Definition
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SISAQOL	Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data
SO	secondary outcome
SOC	system organ class
TEAE	treatment-emergent adverse event
TIMP	Trial Issue Management Plan
TFFS	treatment failure-free survival
TTNT	time to initiation of new anticancer therapy
XML	extensible markup language

Leo Document ID = e541c922-2ceb-47af-be6f-800d0c2b07e3

Approver: PPD

Approval Date & Time: 16-Dec-2021 16:35:01 GMT

Signature meaning: Approved