NOVARTIS

Clinical Development

Dabrafenib (Tafinlar), Trametinib (Mekinist)

CDRB436J12301 / NCT04940052

A randomized, double-blind, placebo-controlled phase III study to evaluate the efficacy and safety of dabrafenib plus trametinib in previously treated patients with locally advanced or metastatic, radio-active iodine refractory BRAFV600E mutation-positive Differentiated Thyroid Cancer

Statistical Analysis Plan (SAP)

Document type:	SAP Documentation
Document status:	Final Amendment 1
Release date:	22-Jan-2025
Number of pages:	57

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
30-Aug-2021	Prior to FPFV	Creation of final version	N/A – First version	NA
22-Jan-2025	Prior to DB lock	Creation of Amendment 1	Adding subgroup analysis for Chinese patients (mainland China) to some analyses	Various sections
			Adding that a BOR of complete or partial response needs to be confirmed	2.6.1 Key secondary endpoints
			Adding summary of Disease Control Rate	2.6.5 Supportive analyses
			Adding crossover phase baseline and treatment period definitions	2.1.2 Crossover phase
			Adding info about analyses based on the Crossover Population Set	Various sections (2.3, 2.4.1, 2.8)
			Adding definition of average daily dose	2.4.1 Study treatment / compliance
			Removal of corticosteroid summary; added a note regarding post anti-neoplastic therapies crossover patients	2.4.2 Prior, concomitant and post therapies
			Update to visit window definitions	2.1.1 General definitions
			Removed analysis of time to first occurrence of AESI	2.8.1.1 Adverse events of special interest / grouping of AEs)
			Update criteria for liver abnormalities, based on latest Novartis guidance	2.8.3 Laboratory data
			Adding analysis of ECOG PS	2.8.4.3 ECOG PS
			Update to list of serous retinopathy events, based on MedDRA 27.1	2.8.5 Serous retinopathy ocular events
			[redacted]	
			Clarified computation details for ORR in case sampling assumptions are not met.	5.4.2 Analysis supporting secondary objective(s)
			Edits made to correct typo or improve clarity/consistency of the document.	Various sections

Table of contents

List of tables

List of figures

List of abbreviations

AE	Adverse Event
AESI	Adverse Events of Special Interest
ALB	Albumin
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Classification
BIRC	Blinded Independent Review Committee
BMI	Body Mass Index
BOR	Best Overall Response
CI	Confidence Interval
CR	Complete Response
CRS	Case Retrieval Strategy
CSR	Clinical Study Report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DAR	Dose Administration Record
DI	Dose Intensity
DMC	Data Monitoring Committee
DOR	Duration of Response
ECOG PS	Eastern Cooperative Oncology Group Performance Status
ECG	Electrocardiogram
ECHO	Echocardiogram
eCRF	Electronic Case Report Form
EOT	End of Treatment
[redacted]	[redacted]
FAS	Full Analysis Set
HR	Hazard Ratio
KM	Kaplan-Meier
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
MUGA	Multigated Acquisition
NCI	National Cancer Institute
NMQ	Novartis MedDRA Query
ORR	Overall Response Rate
OS	Overall Survival
[redacted]	[redacted]
PD	Pharmacodynamics
PDI	Planned Dose Intensity
PFS	Progression-Free Survival
[redacted]	[redacted]
PR	Partial Response
[redacted]	[redacted]
PT	Preferred Term
[redacted]	[redacted]

RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SD	Stable Disease
SMQ	Standardized MedDRA Query
SOC	System Organ Class
TA	Tumor Assessment
TBL	Total Bilirubin
ULN	Upper Limit of Normal
UNK	Unknown
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes all planned analyses of the clinical study report (CSR) of study CDRB436J12301, A randomized, double-blind, placebo-controlled phase III study to evaluate the efficacy and safety of dabrafenib plus trametinib in previously treated patients with locally advanced or metastatic, radio-active iodine refractory BRAF V600E mutation-positive differentiated thyroid cancer.

The content of this SAP is based on CDRB436J12301 protocol amendment 03. All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock of the study data.

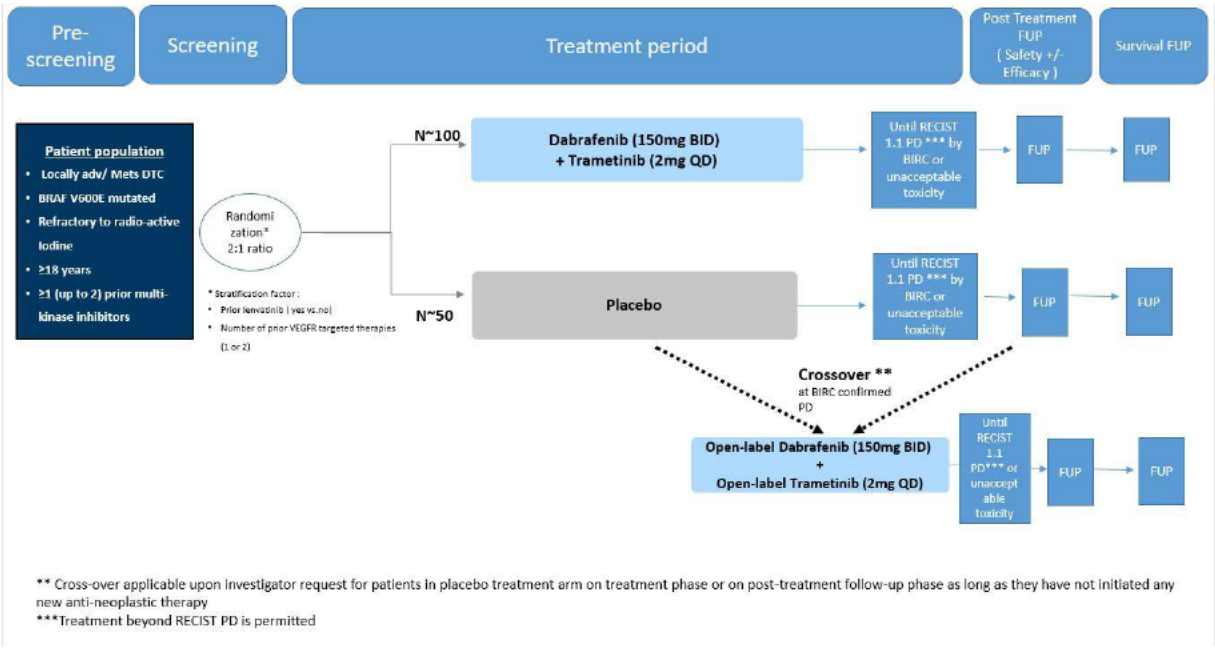
1.1 Study design

This is a global, multicenter, randomized, double-blind, placebo-controlled phase III study to evaluate the efficacy and safety of dabrafenib plus trametinib in adult patients with locally advanced or metastatic BRAF V600E mutation-positive, differentiated thyroid carcinoma who are refractory to radioactive iodine and have progressed following prior VEGFR targeted therapy. After eligibility assessment, approximately 150 patients will be randomized in a 2:1 ratio to either dabrafenib plus trametinib or placebo. Patients will be stratified by number of prior VEGFR targeted therapy (1 versus 2) and prior lenvatinib treatment (yes versus no). For more details on the design, refer to the protocol.

The scientific objective guiding the primary estimand is based on the PFS as per BIRC assessment using RECIST 1.1 criteria. No independent Data Monitoring Committee (DMC) or interim analyses are planned. The primary analysis will be performed when all patients have completed approximately 16 weeks of follow-up or have discontinued before (for additional details please refer to Section 2.5.1). At this time it is expected that approximately 95 PFS events will have been reported as per BIRC assessment.

Refer to Figure 1-1 for an overview of the study design.

Figure 1-1 Study Design



** Cross-over applicable upon investigator request for patients in placebo treatment arm on treatment phase or on post-treatment follow-up phase as long as they have not initiated any new anti-neoplastic therapy
***Treatment beyond RECIST PD is permitted

1.2 Study objectives, endpoints and estimands

Objectives and related endpoints are described in Table 1-1 below.

Table 1-1 Objectives and related endpoints

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• To compare PFS between dabrafenib plus trametinib versus placebo	• PFS based on BIRC assessment using RECIST 1.1 criteria See Section 1.2.1 for Primary Estimand
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**

• **Key efficacy secondary objectives:**
  • to compare ORR of dabrafenib plus trametinib versus placebo
  • to compare OS of dabrafenib plus trametinib versus placebo
• **Other efficacy secondary objective:**
  • to evaluate DOR of dabrafenib plus trametinib versus placebo
	• ORR as per BIRC assessment using RECIST 1.1 criteria
	• OS including all deaths from any cause
	See Section 1.2.2 for Secondary Estimands.
	• DOR by BIRC assessment using RECIST v1.1 criteria

• To characterize the safety and tolerability of dabrafenib and trametinib	• Incidence and severity of AEs and SAEs, including changes in laboratory values, ECOG PS (performance status), and vital signs.
• To quantify trametinib associated serous retinopathy ocular events	• Incidence, type and severity of trametinib associated serous retinopathy ocular events.
[redacted]	

1.2.1 Primary estimand(s)

The clinical question of interest is: what is the relative treatment effect of dabrafenib plus trametinib versus placebo in prolonging the time to progression or death had the new antineoplastic therapies not occurred regardless of treatment discontinuation and any pandemic related events in patients with previously treated BRAFV600E mutation positive DTC.

The justification for targeting this treatment effect is that we wish to estimate the relative effect of the two treatments in the absence of potential confounding effect of any new antineoplastic therapy that is not part of the assigned treatment strategy and that may potentially occur more frequently in the placebo arm than in the dabrafenib plus trametinib arm.

The primary estimand is described by the following attributes:

- Population: adult patients with locally advanced or metastatic DTC, RAI-r, with BRAFV600E mutation and who have progressed following prior VEGFR-targeted therapies (no more than 2).
- Primary variable: progression-free survival as assessed by BIRC, using RECIST v1.1, defined as the time from randomization to disease progression or death due to any cause, whichever occurs first.
- Treatment of interest: dabrafenib plus trametinib or placebo. Further details about the treatments are provided in Section 2.4.
- Handling of the remaining intercurrent events:
  - Treatment discontinuation due to any reason will be handled using treatment policy strategy since all PFS events will be considered as an event irrespective of the study treatment discontinuation reasons.
  - Any unforeseen intercurrent events (e.g., pandemic related events, such as COVID-19) will be handled using treatment policy strategy.

  - Initiation of any new antineoplastic therapy started before observing any PFS event will be handled using the hypothetical strategy, i.e., as if new anticancer therapy had not been available. New antineoplastic therapy has the potential to confound the interpretation of effect of the treatment strategy, especially if this occurs more frequently in placebo arm versus dabrafenib plus trametinib arm.
- The summary measure is the hazard ratio (HR) for PFS between the two treatment arms, estimated using a stratified Cox proportional hazard model.

Supplementary estimands to the primary estimand are defined in Section 12 in the protocol.

1.2.2 Secondary estimand(s)

The following two key secondary estimands for efficacy are considered.

The secondary clinical question of interest isrelated to the treatment effect based on overall response rate between dabrafenib plus trametinib and placebo, regardless of treatment discontinuation and any unforeseen pandemic events (e.g., COVID-19), for patients with BRAFV600E mutation-positive advanced/metastatic DTC.

The secondary estimand linked to this secondary question is described by the following attributes:

- Population: adult patients with locally advanced or metastatic DTC, refractory to radioactive iodine, with BRAFV600E mutation and who have progressed following prior VEGFR-targeted therapies (no more than 2).
- Primary variable: best overall response, defined as the best response recorded from the start of the treatment up to 30 days after the last dose of study treatment or disease progression as per BIRC using RECIST 1.1 criteria, whichever occurs first, with responses after the use of new antineoplastic therapy or after the start of open-label dabrafenib and trametinib treatment following crossover for placebo patients considered as non-responses.
- Treatment of interest: dabrafenib plus trametinib or placebo.
- Handling of the remaining intercurrent events:
  - Treatment discontinuation due to any reason will be handled using treatment policy strategy since all response assessments occurring after treatment discontinuation and during the 30-days post-treatment follow-up period will be considered for BOR derivation irrespective of the study treatment discontinuation reasons.
  - Any unforeseen intercurrent events (e.g., pandemic related events, such as COVID-19) will be handled using treatment policy strategy.
- The summary measure is the difference in ORR of the two treatments (defined as the proportion of patients with confirmed BOR of complete response (CR) or partial response (PR) based on BIRC per RECIST 1.1) and its 95% confidence interval calculated using exact method.

One other secondary clinical question of interest is the relative treatment effect of dabrafenib plus trametinib versus placebo in prolonging the survival time, regardless of treatment discontinuation, new antineoplastic therapies, any pandemic related events and crossover, for patients with BRAFV600E mutation-positive advanced or metastatic DTC.

The secondary estimand is described by the following attributes:

- Population: adult patients with locally advanced or metastatic DTC, refractory to radioactive iodine, with BRAFV600E mutation and who have progressed following prior VEGFR-targeted therapies (no more than 2).
- Primary variable: overall survival, defined as the time from randomization to death due to any cause.
- Treatment of interest: dabrafenib plus trametinib or placebo with or without any new antineoplastic therapy received post randomization as needed.
- Handling of the remaining intercurrent events:
  - Treatment discontinuation due to any reason will be handled using treatment policy strategy since all deaths will be considered as an event irrespective of the study treatment discontinuation reasons.
  - Any unforeseen intercurrent events (e.g., pandemic related events, such as COVID-19) will be handled using treatment policy strategy.
  - Crossover of patients from placebo to dabrafenib plus trametinib will be handled using treatment policy strategy: all deaths will be considered as an event irrespective of the crossover.
- The summary measure is the hazard ratio (HR) for OS between the two treatment arms, estimated using stratified Cox proportional hazard model.

The following secondary estimand for safety is considered in the scope of a post-marketing requirement: what is the incidence of trametinib associated serous retinopathy ocular events while on-treatment, regardless of new antineoplastic therapies and any pandemic related events.

The secondary estimand is described by the following attributes:

- Population: adult patients with locally advanced or metastatic DTC, refractory to radioactive iodine, with BRAFV600E mutation and who have progressed following prior VEGFR-targeted therapies (no more than 2), who have received at least one dose of the study treatment and have at least two on-treatment OCT assessments.
- Primary variable: occurrence of serous retinopathy grouping event, reported in the AEs CRF (case report form) page, on-treatment, occurring from the start of study treatment and up to 30 days after last dose of study treatment, using specific case retrieval strategy for ocular events, listing the appropriate preferred terms falling into the definition of a serous retinopathy.
- Treatment of interest: dabrafenib plus trametinib.
- Handling of the remaining intercurrent events:
  - Treatment discontinuation of trametinib or dabrafenib at any time due to any reason will be handled using treatment policy strategy since all serous retinopathy events occurring on-treatment or up to 30 days after the last dose of study treatment will be considered irrespective of the trametinib or dabrafenib discontinuation reasons.
  - Any unforeseen intercurrent events (e.g., pandemic related events, such as COVID-19) will be handled using treatment policy strategy.
  - Any new antineoplastic therapies received during treatment or during the 30-days post treatment follow-up period will be handled using treatment policy strategy: all serous

retinopathy events will be considered regardless of the initiation of a new antineoplastic therapy

- The summary measure is the proportion of patients with at least one serous retinopathy grouping event occurring on-treatment, over the Ocular Event Evaluable Set.

One supplementary analysis of this safety secondary estimand will be done by changing the population of interest and the summary measure attributes, using the patients included in the Safety Set.

2 Statistical methods

2.1 Data analysis general information

All analysis will be performed by Novartis. SAS version 9.4 or later software will be used to perform all data analyses and to generate tables, figures and listings.

Data included in the analysis

For each of the analyses, all statistical analyses will be performed using all data collected in the database up to the data cut-off date. All data with an assessment date or event start date (e.g., vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.

All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'ongoing'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these cases, the end date will not be imputed and therefore will not appear in the listings.

There is one primary analysis planned for the primary efficacy endpoint (PFS), and one final analysis may be performed for the key secondary endpoints (OS, ORR).

The primary analysis will be performed when all patients have completed approximately 16 weeks of follow-up or have discontinued before (for additional details please refer to Section 2.5.1). At this time, it is expected that approximately 95 PFS events will have been reported as per BIRC assessment, and the primary clinical study report (CSR) will be produced. After the primary analysis of PFS, the study will remain open provided the PFS demonstrates treatment benefit. Patients still being followed on the study after the primary analysis time point will continue as per the schedule of assessments.

Analysis for OS will be performed at the time of the primary analysis for PFS, provided PFS and ORR are significant, at which point a total of approximately 39 deaths are expected (around 31 months from first patient randomized).

The analysis cut-off date for the final OS analysis will be established when all patients have been followed for at least 3 years. If the primary analysis of PFS does not demonstrate treatment benefit, then follow-up for OS will end.

General analysis conventions

Pooling of centers: Unless specified otherwise, data from all study centers will be pooled for the analysis. Due to expected small number of patients enrolled at centers, no center effect will be assessed.

Qualitative data (e.g., gender, race) will be summarized by means of contingency tables by treatment group; a missing category will be included as applicable. Percentages will be calculated using the number of patients in the relevant population or subgroup as the denominator.

Quantitative data (e.g., age, body weight) will be summarized by appropriate descriptive statistics (i.e., mean, standard deviation, median, minimum, and maximum) by treatment group.

2.1.1 General definitions

Investigational drug and study treatment

Investigational drug: dabrafenib or trametinib

Study treatment: dabrafenib/trametinib combination or placebo

Study drug*:* dabrafenib, trametinib, dabrafenib placebo, or trametinib placebo

Open-label study treatment*:* open-label dabrafenib/trametinib combination following crossover from placebo

Date of first administration of study drug

The date of first administration of study drug is defined as the first date when a non-zero dose of study drug is administered and recorded on the Dosage Administration Record (DAR) Electronic Case Report Form (eCRF). The date of first administration of study drug will also be referred as start of study drug.

Date of last administration of study drug

The date of last administration of study drug is defined as the last date when a non-zero dose of study drug is administered and recorded on DAR eCRF. The date of last administration of study drug will also be referred as end of study drug.

Date of first administration of study treatment

The date of first administration of study treatment is derived as the first date when a non-zero dose of any component of study treatment was administered as per the Dosage Administration Record (e)CRF. (Example: if 1st dose of dabrafenib or placebo is administered on 05-Jan-2016, and 1st dose of trametinib is administered on 03-Jan-2016, then the date of first administration of study treatment is on 03-Jan-2016). The date of first administration of study treatment will also be referred as ***start of study treatment***.

Date of last administration of study treatment

The date of last administration of study treatment is derived as the last date when a non-zero dose of any component of study treatment was administered as per Dose Administration Record

(e)CRF. (Example: if the last dabrafenib or placebo dose is administered on 15-Apr-2016, and the last dose of trametinib is administered on 17-Apr-2016, then the date of last administration of study treatment is on 17-Apr-2016).

Study day

The study day describes the day of the event or assessment date, relative to the reference start date. The study day is defined as:

- The date of the event (visit date, onset date of an event, assessment date etc.) – reference start date + 1 if event is on or after the reference date.
- The date of the event (visit date, onset date of an event, assessment date etc.) – reference start date if event precedes the reference date.

The reference start date for safety assessments (e.g., adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, [redacted].) is the start of study treatment.

The reference start date for all other, non-safety assessments (i.e., tumor assessment, survival time, disease progression, tumor response, Eastern Cooperative Oncology Group [ECOG] performance status, [redacted] is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Time unit

A year length is defined as 365.25 days. A month length is 30.4375 days (365.25/12). If duration is reported in months, duration in days will be divided by 30.4375. If duration is reported in years, duration in days will be divided by 365.25.

Baseline

For efficacy evaluations, the last non-missing assessment, including unscheduled assessments on or before the date of randomization is defined as “baseline” value or “baseline” assessment. In the context of baseline definition, the efficacy evaluations also include [redacted] performance status.

For safety evaluations, the last available assessment on or before the date of start of study treatment is defined as ‘‘baseline” assessment. For cases where time of assessment and time of treatment start is captured (e.g. pre-dose ECG, laboratory assessments), the last available assessment before the treatment start date/time is used for baseline.

In rare cases where multiple measurements meet the baseline definition, with no further flag or label that can identify the chronological order, then the following rule should be applied: If values are from central and local laboratories, the value from central assessment should be considered as baseline. If multiple values are from the same laboratory (local or central) or collected for ECGs or vital signs, then the last value should be considered as baseline.

If patients have no value as defined above, the baseline result will be missing.

On-treatment assessment/event and observation periods

The overall observation period will be divided into three mutually exclusive segments:

1. **Pre-treatment period**: from day of patient's informed consent to the day before first administration of the study treatment
2. **On-treatment period**: from day of first administration of the study treatment to the earlier of
   - 30 days after date of last administration of the study treatment
   - day prior to the start of open-label dabrafenib and trametinib treatment following crossover from placebo
3. **Post-treatment period**: starting at day 31 after last administration of study treatment or at the first day of open-label dabrafenib and trametinib treatment.

For cases where time of assessment and time of treatment start/stop is captured (e.g., ECG's, laboratory assessments), the last available assessment before the treatment period start/stop date/time will be used.

If dates are incomplete in a way that clear assignment to pre-, on-, post-treatment period cannot be made, then the respective data will be assigned to the on-treatment period. Refer to Section 5.1.2 for imputation rules concerning AE start and stop dates.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g., change from baseline summaries). In addition, a separate summary for deaths including on-treatment and post-treatment deaths will be provided. Summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date within the on-treatment period (treatment-emergent AEs). However, all safety data (including those from the post-treatment period) will be listed and those collected during the pre-treatment and post-treatment period will be flagged.

Windows for multiple assessments

In order to summarize [redacted] measures, performance status (ECOG), physical exam, vital sign, ECG, ECHO, laboratory, [redacted] collected over time (including unscheduled visits), the assessments will be time slotted. The following general rule will be applied in creating the assessment windows: If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If 2 assessments within a time window are equidistant from the target date, then the earlier of the 2 assessments will be used. For multiple assessments on the same date, the worst case will be used. In the situation that multiple assessments are all normal, the average will be taken. Data from all assessments (scheduled and unscheduled), including multiple assessments, will be listed. Assessments included in the EOT assessment will also be available for inclusion in the other time assessment windows. Please see the Protocol Table 8-1 for details.

Table 2-1 Time windows for ECOG PS assessments

Assessment	Target day of assessment	Time Interval
Baseline	1	≤ Day 1
Week 4	28	Day 2 to 41

Every 4 weeks through week 52 Week 4*k (with k=2,…13)	d=4*k*7	Day 4*k*7-14 to4*k*7+13
Week 56	392	Day 378 to 433
Every 12 weeks thereafter Week 56+12*k (with k=1,2,…)	d=(56+12*k)*7	Day (56+12*k)*7-42 to (56+12*k)*7+41
End of Treatment		Assessment taken at the end of treatment visit
Study Day 1 = randomization date		

Table 2-2 Time windows for laboratory assessments

Assessment	Target day of assessment	Time Interval
Baseline	1	≤ Day 1
Week 4	28	Day 2 to 41
Every 4 weeks through week 52 Week 4*k (with k=2,…13)	d=4*k*7	Day 4*k*7-14 to 4*k*7+13
Week 56	392	Day 378 to 433
Every 12 weeks thereafter Week 56+12*k (with k=1,2,…)	d=(56+12*k)*7	Day (56+12*k)*7-42 to (56+12*k)*7+41
End of Treatment		Assessment taken at the end of treatment visit
Study Day 1 = the date of start of study treatment		

Table 2-3 Time windows for ECG and LVEF (ECHO or MUGA) assessments

Assessment	Target day of assessment	Time Interval
Baseline	1	≤ Day 1
Week 4	28	Day 2 to 69
Week 16	112	Day 70 to 153
Week 28	196	Day 154 to 237
Week 40	280	Day 238 to 321
Week 52	364	Day 322 to 377
Week 56	392	Day 378 to 433
Every 12 weeks thereafter Week 56+12*k (with k=1,2,…)	d=(56+12*k)*7 476	Day (56+12*k)*7-42 to (56+12*k)*7+41
End of Treatment		Assessment taken at the end of treatment visit
Study Day 1 = the date of start of study treatment		

Time windows will be defined for descriptive summary of [redacted] data by visit. If more than one assessment is available in the same time window, the assessment closest to the planned date will be considered. If two assessments are obtained with the same time difference compared to the scheduled visit day, the assessment obtained prior to visit will be considered. Data obtained at the end of treatment will be classified as other assessment in the corresponding time window. The end of treatment assessment will be included if collected within 30 days of the last dose intake.

[redacted]

Last contact date

The last contact date will be derived for patients not known to have died at the analysis cut-off using the last complete date among the following:

Table 2-5 Last contact date data sources

Source data	Conditions
Date of Randomization	No condition
Last contact date/last date patient was known to be alive from Survival Follow-up page	Patient status is reported to be alive, lost to follow-up or unknown.
Start/End dates from further antineoplastic therapy	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
End of treatment date from end of treatment page	No condition.
Tumor (RECIST) assessment date	Evaluation is marked as 'done'.
Verification for treatment beyond RECIST1.1 PD	At least one non-missing parameter value.
Laboratory/[redacted]	Sample collection marked as 'done'.
Vital signs date	At least one non-missing parameter value
Performance Status date	Non-missing performance status
Start/End dates of AE	Non-missing verbatim term
[redacted]	[redacted]
[redacted]	[redacted]

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date. The cut-off date will not be used for last contact date, unless the patient was seen or contacted on that date. No date post cut-off date will be used. Completely imputed dates (e.g., the analysis cut-off date programmatically imputed to replace the missing end date of a dose administration record) will not be used to derive the last contact date. Partial date imputation is allowed for event (death)/censoring coming from the 'Survival information' eCRF.

The last contact date will be used for censoring of patients in the analysis of overall survival.

2.1.2 Crossover phase

Baseline

Baseline is defined as the most recent non-missing value on or prior to the day of the first dose of study treatment (dabrafenib plus trametinib) on the crossover treatment period. Baseline values will be established prior to the start of the crossover phase.

Response will be determined separately for the randomized, double-blind, placebo-controlled phase and the crossover phase. Baseline lesion assessments will be re-established prior to initiation of crossover therapy and response will be calculated based on the appropriate baseline for each respective phase.

On-treatment assessment/event and observation periods

For adverse event reporting the overall observation period will be divided into three mutually exclusive segments:

1. pre-treatment period: up to 90 days prior to the first dose on the crossover phase
2. on-treatment period: from date of first administration of crossover study treatment to 30 days after date of last actual administration of study treatment (including start and stop date)
3. post-treatment period: starting at day 30+1 after last administration of crossover study treatment.

The reference date for both efficacy and safety measures in the Crossover phase is the date of first dose of dabrafenib plus trametinib on the crossover treatment period. Refer to Section 2.1.1 for time windows for the Crossover phase relative to this reference date. Note: the end of treatment visit for the randomized placebo-controlled phase and the start of crossover treatment visit might happen on the same day.

Separate tables and listings will be used for the randomized, double-blind, placebo-controlled phase and the crossover phase.

Note: Analyses specified as for the randomized, double-blind, placebo-controlled phase will only include data prior to crossover except for analyses of overall survival (OS). Similarly analyses specified for the crossover phase will only use data from after the date of crossover.

2.2 Analysis sets

Full Analysis Set

The Full Analysis Set (FAS) comprises all patients to whom study treatment has been assigned by randomization. According to the intent to treat principle, patients will be analyzed according to the treatment and strata they have been assigned to during the randomization procedure. This population will be the primary population for efficacy analyses.

Per Protocol Set

Not applicable.

Safety Set

The Safety Set includes all patients who received at least one dose of any component of the study treatment. Patients will be analyzed according to the study treatment received, where treatment received is defined as the randomized/assigned treatment if the patient took at least one dose of that treatment or the first treatment received if the randomized/assigned treatment was never received.

Ocular Event Evaluable Set

The Ocular Event Evaluable Set comprises all patients from the Safety Set who have one baseline and at least two on-treatment optical coherence tomography (OCT) assessments.

[redacted]

- [redacted]
- [redacted].
- [redacted]

[redacted]

Crossover Population Set

The Crossover Population Set may be used to summarize the analyses performed on data collected after the crossover, if applicable: it includes all patients who received at least one dose of any component of the open-label study treatment.

2.2.1 Subgroup of interest

Efficacy

The primary efficacy and key secondary endpoint (PFS, ORR and OS) will be summarized by the following subgroups ***to examine the homogeneity of treatment effect*** provided that the primary efficacy analysis based on the FAS is statistically significant:

- Gender (male vs female)
- Age (<65 vs ≥ 65 years)
- Region (China vs rest of the world)
- Type of prior treatment received (lenvatinib vs others)
- Number of prior VEGFR targeted therapies (1 vs 2)
- Baseline TSH (≤0.5 vs >0.5 mIU per liter)
- Baseline thyroglobulin (≤10 vs. >10 ng/ml)

No formal statistical test of hypotheses will be performed for the subgroups, only point estimates of the treatment effect and 95% confidence intervals will be provided. The objective of the efficacy subgroup analysis is to demonstrate homogeneity of treatment effect in the above subgroups. Summary tables will only be generated if at least 10% of patients or 10 patients are present in each subgroup.

2.3 Patient disposition, demographics and other baseline characteristics

The Full Analysis Set (FAS) will be used for all baseline (including disease characteristics) and demographic summaries and listings unless otherwise specified. Summaries will be reported by treatment group and for all patients, and listings will be reported by treatment group to assess baseline comparability. No inferential statistics will be provided. Selected key baseline characteristics will be summarized based on the Crossover Population Set.

2.3.1 Demographics and other baseline characteristics

Basic demographic and background data

All demographic and baseline disease characteristics data will be summarized and listed by treatment arm. Categorical data (e.g., gender, age groups: <65 and ≥ 65 years, region: China vs rest of the world, type of prior treatment received; lenvatinib vs others, number of prior VEGFR targeted therapies: 1 vs 2, baseline TSH: ≤0.5 vs >0.5 mIU per liter, baseline thyroglobulin: ≤10 vs. >10 ng/ml) will be summarized by frequency counts and percentages; the number and percentage of patients with missing data will be provided. Continuous data (e.g., age, weight, height, body mass index) will be summarized by descriptive statistics (N, mean, median, standard deviation, minimum and maximum). BMI (kg/m^2) will be calculated as weight [kg] / (height$[m]^2$) using weight at screening.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

Baseline stratification factors

The number and percentage (%) of patients in each stratum (prior lenvatinib treatment (yes vs no), number of prior VEGFR targeted therapies (1 vs 2)) based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS. Discordances between the stratum recorded in IRT at the time of randomization and the actual stratum recorded in the clinical database through the data collected on eCRF will be cross-tabulated and listed.

Diagnosis and extent of cancer

Summary statistics will be tabulated for diagnosis and extent of cancer. This analysis will include the following: primary site of cancer, details of tumor histology/cytology, histological grade, stage at initial diagnosis, time since initial diagnosis, time from initial diagnosis to first recurrence/progression (in months), time since most recent relapse/progression to randomization (in months), stage at time of study entry, presence/absence of target and non-target lesions, number and type of metastatic sites involved. Note: Presence/absence of target and non-target lesions will be based on the data collected on the RECIST target/non-target lesion assessment eCRF pages. Metastatic sites will be based on diagnosis page.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

Medical history

Medical history and ongoing conditions, including cancer-related conditions and symptoms entered on eCRF will be summarized and listed by treatment group. Separate summaries will

be presented for ongoing and historical medical conditions. The summaries will be presented by primary system organ class (SOC), preferred term (PT) and treatment arm. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

[redacted]

2.3.2 Patient disposition

Enrollment by country and center will be summarized for all screened patients and also by treatment group using the FAS. The number (%) of randomized patients included in the FAS will be presented overall and by treatment group. The number (%) of screened and not-randomized patients and the reasons for screen failures will also be displayed. The number (%) of patients in the FAS who are still on treatment, who discontinued the study phases and the reason for discontinuation will be presented overall and by treatment group.

The following summaries will be provided (with % based on the total number of FAS patients):

- Number (%) of patients who were randomized (based on data from IRT system);
- Number (%) of patients who were randomized but not treated (based on 'DAR' eCRF page not completed for any study treatment component);
- Primary reason for not being treated (based on 'End of Treatment Disposition' eCRF page);
- Number (%) of patients who were treated (based on 'DAR' eCRF pages of each study treatment component completed with non-zero dose administered);
- Number (%) of patients who are still on-treatment (based on the 'End of Treatment Disposition' page not completed);
- Number (%) of patients who discontinued the study treatment phase (based on the 'End of Treatment Disposition' page);
- Primary reason for study treatment phase discontinuation (based on the 'End of Treatment Disposition' page);
- Number (%) of patients who have entered the post-treatment follow-up (based on the 'End of Treatment Disposition' page);
- Number (%) of patients who have discontinued from the post-treatment follow-up (based on the 'End of Post Treatment Follow-up Disposition' page);
- Reasons for discontinuation from the post-treatment follow-up (based on 'End of Post Treatment Follow-up Disposition' page);
- Number (%) of patients who have entered the survival follow-up (based on the 'End of Treatment Disposition' or 'End of Post Treatment Follow-up Disposition' page).

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

Protocol deviations

The number (%) of patients in the FAS with any protocol deviations will be tabulated by deviation category (as specified in the study Data Handling Plan) overall and by treatment group. All protocol deviations will be listed. Specific Protocol Deviation categories will be assigned to important deviations related to COVID-19 and will be summarized and listed separately.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

Analysis sets

The number and percentages (based on the total number of FAS patients) of patients in each analysis set (defined in Section 2.2) will be summarized.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

Duration of exposure, actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized by treatment group, separately for each component of study treatment (dabrafenib and trametinib). The duration of treatment will also be presented for each combination arm. Duration of exposure will be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of subjects in each interval. The number (%) of subjects who have dose reductions or interruptions, and the reasons, will be summarized by treatment group.

Subject level listings of all doses administered on treatment along with dose change reasons will be produced.

The Safety Set will be used for all summaries and listings of study treatment. The summaries described above will be also provided using the Crossover Population Set.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

Duration of exposure to study treatment

Duration of exposure to study treatment is considered by taking into account the duration of exposure to any combination partner:

Duration of exposure to study treatment (days) = (last date of exposure to study treatment) - (date of first administration of study treatment) + 1.

The last date of exposure to study treatment is the latest of the last dates of exposure to any combination partner (see Table 2-6).

Summary of duration of exposure of study treatment in appropriate time units will include categorical summaries and continuous summaries (i.e., mean, standard deviation etc.) using appropriate units of time.

Duration of exposure to combination partner

Duration of exposure to study drug (days) = (last date of exposure to study drug) - (date of first administration of study drug) + 1.

Refer to Table 2-6 for definitions of the last date of exposure to dabrafenib and trametinib. Study drug is defined as dabrafenib, trametinib, dabrafenib placebo, or trametinib placebo.

Table 2-6 Definition of last date of exposure of study drug

Scenario	Definition of last date of exposure of study drug	Example
Dabrafenib, Trametinib	Date of last administration of a non-zero dose of the study drug.	Example 1: A patient had a permanent discontinuation of the study drug on 06Jan2016 after being put on a temporary interruption since 01Jan2016. In this case the last date of exposure is 31Dec2015.

Summary of duration of exposure of dabrafenib and trametinib or placebo will include categorical summaries based on 28 day intervals and using descriptive statistics (mean, standard deviation, etc.).

Cumulative dose and average daily dose

Cumulative dose of a study treatment is defined as the total dose given during the study treatment and will be summarized for each of the study treatment components (dabrafenib, trametinib, dabrafenib placebo, trametinib placebo). Average daily dose is defined as [Cumulative dose (dosing unit) / Number of dosing days]; drug free days are not counted as dosing days.

The **planned cumulative dose** for a study treatment component refers to the total planned dose as per the protocol up to the last date of investigational drug administration. The planned cumulative dose will not be summarized/listed. It will be used for relative dose intensity calculations.

The **actual cumulative dose** refers to the total actual dose administered, over the duration for which the patient is on the study treatment as documented in the Dose Administration Record eCRF.

For patients who did not take any drug the cumulative dose is by definition equal to zero.

For continuous dosing, the actual cumulative dose is the sum of the non-zero doses recorded over the dosing period and the planned cumulative dose is the planned starting dose summed over the same dosing period.

For intermittent dosing, the actual cumulative dose should be defined based on the days when the patient is assumed to have taken a non-zero dose during dosing periods.

Dose intensity and relative dose intensity

Dose intensity (DI) for patients with non-zero duration of exposure is defined as follows:

DI (mg */ unit of time*) = Actual Cumulative dose (mg) / Duration of exposure to study treatment (*unit of time*).

For patients who did not take any drug the DI is by definition equal to zero.

Planned dose intensity (PDI) is defined as follows:

PDI (mg */ unit of time)* = Planned Cumulative dose (mg) / Duration of exposure (*unit of time*).

Relative dose intensity (RDI) is defined as follows:

RDI (%) = [DI (mg */ unit of time*) / PDI (mg */ unit of time*)] x 100.

For dabrafenib and trametinib, the unit of time will be 1 day.

DI and RDI will be summarized for combination studies separately for each of the study treatment components, but using the duration of exposure of each of the components.

Table 2-7 Examples of dabrafenib (or dabrafenib placebo) dose administration and exposure

DAR record number	Start/End Date	Dose Prescribed (mg), frequency	Dose Administered (mg) [total daily]	Dose Changes, Dose Interruption?	Dose Permanently Discontinued	Reason
1	01Jan2016 / 05Jan2016	150 mg BID	300	No	No	
2	06Jan2016 / 03Feb2016	150 mg BID	200	Yes	No	AE
3	04Feb2016 / 25Feb2016	150 mg BID	300	No	No	
Duration of exposure (days) = 25Feb2016 – 01Jan2016 + 1 = 56 days Planned cumulative dose (for 56 days) = 300*56 days = 16800 mg Actual cumulative dose = 300*5 + 200*29 + 300*22 = 13900 mg Dose intensity = 13900 mg / 56 days = 248.21 mg/day Planned dose intensity = 16800 mg / 56 days = 300 mg/day Relative dose intensity = DI / PDI = (248.21 mg/day) / (300 mg/day) = 83%						

Table 2-8 Examples of trametinib (or trametinib placebo) dose administration and exposure

DAR record number	Start/End Date	Dose Prescribed (mg), frequency	Dose Administered (mg) [total daily]	Dose Changes, Dose Interruption?	Dose Permanently Discontinued	Reason
1	01Jan2016 / 10Jan2016	2 QD	2	No	No	
2	11Jan2016 / 15Jan2016	2 QD	0	Yes	No	AE
3	16Jan2016 / 25Feb2016	1 QD	1	No	No	AE
Duration of exposure = 25Feb2016 – 01Jan2016 + 1 = 56 days Planned cumulative dose (for 56 days) = 2*56 days = 112 mg Actual cumulative dose = 2*10 + 0*5 + 1*41 = 61 mg Dose intensity = 61 mg / 56 days = 1.09 mg/day						

Planned dose intensity = 112 mg / 56 days = 2 mg/day
Relative dose intensity = DI / PDI = (1.09 mg/day) / (2 mg/day) = 54%

Dose reductions, interruptions or permanent discontinuations

The number of patients who have dose reductions, permanent discontinuations, or interruptions, and the reasons, will be summarized separately for each of the study treatment components.

'Dose interrupted', and 'Dose permanently discontinued' fields from the Dosage Administration Record CRF pages (DAR) will be used to determine the dose reductions, dose interruptions, and permanent discontinuations, respectively.

The corresponding fields 'Reason for dose change/dose interrupted', and 'Reason for permanent discontinuation' will be used to summarize the reasons.

A dose change is either 'change in prescribed dose level' or 'dosing error' where actual dose administered/total daily dose is different from the prescribed dose.

For the purpose of summarizing interruptions and reasons, in case multiple entries for interruption that are entered on consecutive days/dose administrations with different reasons, separate interruptions will be counted. However, if the reason is the same for multiple entries on consecutive days/dose administrations, then it will be counted as one interruption.

Reduction: a dose change where the prescribed dose level is lower than the previous prescribed dose level or where the actual dose administered/total daily dose is lower than the calculated dose amount based on the prescribed dose. Therefore any dose change to correct a dosing error will not be considered a dose reduction. Only dose change is collected in the CRF, number of reductions will be derived programmatically based on the change and the direction of the change.

Treatment beyond RECIST progression

The number of patients who continue treatment beyond RECIST1.1 progression according to local investigators assessment and confirmed by BIRC based on protocol specified criteria will be summarized. It includes all patients who received any study treatment (i.e., at least one dose of any component of the study treatment) after RECIST 1.1 progression as confirmed by BIRC. Those patients will be identified using the field "Will the subject continue treatment beyond disease progression as per RECIST1.1?" on the Verification for Treatment beyond RECIST1.1 PD CRF pages.

2.4.2 Prior, concomitant and post therapies

Prior anti-neoplastic therapy

The number and percentage of patients who received any prior anti-neoplastic medications, prior anti-neoplastic radiotherapy or prior anti-neoplastic surgery, or other therapies not defined by the previous categories, will be summarized by treatment arm. Prior anti-neoplastic medications will be summarized by therapy type (e.g., chemotherapy, hormonal therapy), setting (e.g., adjuvant, metastatic) and also by the lowest Anatomical Therapeutic Classification (ATC) class, preferred term and treatment. Summaries will include total number of regimens, best response and time from last treatment to progression for the last therapy. The medication therapy type of any combination therapy will be classified based on the following order:

immunotherapy, chemotherapy, biologic therapy (other than immunotherapy), targeted therapy, hormonal therapy. For radiotherapy, time since last radiotherapy, locations and setting of last therapy will be summarized. For prior surgery, time since last surgery and procedure will be summarized.

Separate listings will be produced for prior anti-neoplastic medications, radiotherapy, and surgery.

Anti-neoplastic medications will be coded using the World Health Organization (WHO) Drug Dictionary (WHO-DD); anti-neoplastic surgery will be coded using MedDRA. Details regarding MedDRA and WHO-DD version will be included in the footnote in the tables/listings.

The above analyses will be performed using the FAS.

Post treatment anti-cancer therapy

Anti-neoplastic therapies since discontinuation of study treatment will be listed and summarized by ATC class, preferred term, overall and by treatment group by means of frequency counts and percentages using FAS. A table including reasons for discontinuation and type of progression will be generated. Of note: open-label dabrafenib and trametinib treatment following crossover for placebo patients is reported on a Dose Administration Record CRF page, and is not reported as an anti-neoplastic therapy.

Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) other than the study treatment administered to a patient coinciding with the study treatment period. Concomitant therapy include medications (other than study drugs) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system and summarized by the lowest ATC class and preferred term using frequency counts and percentages. Surgical and medical procedures will be coded using MedDRA and summarized by SOC and preferred term.

All summaries will include:

1. Medications starting on or after the start of study treatment but no later than 30 days after last dose of study treatment and
2. Medications starting prior to start of study treatment and continuing after the start of study treatment.

All reported concomitant therapies will be listed. Any concomitant therapies starting and ending prior to the start of study treatment or starting more than 30 days after the last date of study treatment will be flagged in the listing. The Safety Set will be used for all concomitant medication tables and listings.

Subgroup analyses will be performed for Chinese patients from sites in mainland China for prior anti-neoplastic therapies, concomitant medications and post-treatment anti-cancer therapies.

2.5 Analysis supporting primary objective(s)

The primary objective is to determine whether treatment with dabrafenib in combination with trametinib prolongs progression-free survival (PFS) compared to placebo in previously treated patients with locally advanced or metastatic, radio-active iodine refractory BRAFV600E mutation-positive differentiated thyroid cancer.

2.5.1 Primary endpoint(s)

The primary variable of the primary estimand, PFS, is defined as the time from the date of randomization to the date of the first documented progression according to RECIST 1.1 (see Section 16.1 in the protocol for details) based on BIRC assessment, or death due to any cause.

A minimum of 77 events need to be observed for 90% power, however, sample size needed to be increased to fulfill the ocular-event analysis regulatory requirement. Therefore, the primary analysis will be performed after all randomized patients have been followed for at least 16 weeks (or have discontinued before) at which point it is expected that approximately 95 PFS events have been reported as per BIRC assessment. Data up to cut-off date will be used for the primary analysis.

In the primary analysis, PFS will be censored at the date of the last adequate tumor assessment before the start of a new antineoplastic therapy, if any, if no PFS event is observed prior to the analysis cut-off date using the censoring options from Table 2-9. See Section 2.5.4 for additional details regarding censoring rules and determination of date of last adequate tumor assessment.

2.5.2 Statistical hypothesis, model, and method of analysis

The primary efficacy analysis will be the comparison of the distribution of PFS (based on BIRC assessment of RECIST 1.1 criteria) between the two treatment groups. The following statistical hypothesis will be tested to address the primary efficacy objective:

$$H_{01}: \theta_1 \geq 1 \quad \text{vs.} \quad H_{A1}: \theta_1 < 1$$

where θ_1 is the PFS hazard ratio (dabrafenib and trametinib vs placebo). The analysis to test this hypothesis will consist of a stratified log-rank test at an overall one-sided 2.5% level of significance. Stratification will be based on the randomization stratification factors: prior lenvatinib treatment (yes vs no) and number of prior VEGFR targeted therapies (1 vs 2) as per IRT.

The primary efficacy variable PFS will be analyzed based on the data observed in the FAS up to the cut-off date, according to the treatment group and strata assigned at randomization. The distribution of PFS will be estimated using the Kaplan-Meier method. The results will be plotted graphically by treatment group. The median and 25th and 75th percentiles of PFS along with 95% confidence intervals will be presented by treatment group. A stratified Cox regression will be used to estimate the hazard ratio (HR) of PFS, along with 95% confidence interval using the randomization strata information.

2.5.3 Handling of intercurrent events

The intercurrent events for the primary estimand and handling strategies are described in Section 1.2.1.

2.5.4 Handling of missing values not related to intercurrent event

In the primary analysis, PFS will be censored at the last adequate tumor assessment before the start of a new antineoplastic therapy, if any. If no PFS event is observed prior to the analysis cut-off date, PFS will be censored at the last adequate tumor assessment on or before the analysis cut-off date. Clinical deterioration will not be considered as documented disease progression.

If a PFS event is observed after two or more missing tumor assessments, then PFS will be censored at the last adequate tumor assessment (prior to the first missing assessment and before the PFS event). If a PFS event is observed after the start of a new antineoplastic therapy, then PFS will be censored at the last adequate tumor assessment prior to the start of the new antineoplastic therapy. Patients without any post-baseline tumor assessment and who did not die will be censored at the time of randomization. See Table 2-9 for PFS censoring rules.

Table 2-9 Outcome and event/censoring dates for PFS, TTP, duration of response analysis

Situation		**End Date**[1]	**Outcome**
A	No baseline assessment	Date of randomization/start of treatment[2]	Censored
B	Progression at or before next scheduled assessment	Date of progression	Progressed
C1	Progression or death after **exactly one** missing assessment	Date of progression (or death)	Progressed
C2	Progression or death after **two or more** missing assessments	Date of last adequate assessment	Censored
D	No progression	Date of last adequate assessment	Censored
E	Treatment discontinuation due to 'Disease progression' without documented progression, i.e. clinical progression based on investigator claim	Ignore clinical progression and follow situations above	As per above situations
F	New anticancer therapy given	Date of last adequate assessment prior to new anticancer therapy	Censored
[1]. = End dates are defined as 'date of death', 'date of progression', 'date of last adequate tumor assessment', 'date of next scheduled assessment'. See Protocol Section 16.1.3.2.7 for details			
[2]. =The rare exception to this is if the patient dies no later than the time of the second scheduled assessment as defined in the protocol in which case this is a PFS event at the date of death.			

2.5.5 Sensitivity analyses

As a sensitivity analysis to assess the impact of stratification, the two treatment groups will be compared using the unstratified log-rank test (based on BIRC assessment of RECIST 1.1 criteria). The HR together with the associated 95% confidence interval obtained using the unstratified Cox regression model will also be presented.

If there is a high rate of discrepancy (> 20%) between the strata classifications constructed using the eCRF data and those obtained from the IRT, a sensitivity analysis will be performed in which a stratified Cox regression model will be used to estimate the treatment hazard ratio and the associated 95% confidence intervals based on the eCRF-derived strata. No other inferential statistics will be provided.

2.5.6 Supplementary analyses

As supplementary analyses performed in the FAS, the hazard ratio and 95% confidence interval for PFS based on independent review will be obtained from a stratified and covariate-adjusted Cox model including as potential covariate the following: ECOG ((1 and 2) vs 0), Gender (male vs female), Age (continuous), Region (China vs rest of the world), Type of prior treatment received (lenvatinib vs others), Number of prior VEGFR targeted therapies (1 vs 2), Baseline TSH (≤0.5 vs >0.5 mIU per liter), Baseline thyroglobulin (≤10 vs. >10 ng/ml).

An additional supplementary analysis will handle the intercurrent event of a new antineoplastic therapy started using treatment policy strategy: all PFS events will be considered regardless of the start of a new antineoplastic therapy. The target population, the primary variable, other intercurrent events and summary measure will be the same in this supplementary analysis as for the primary estimand.

If the primary analysis of PFS is statistically significant, subgroup analyses to assess the homogeneity of the treatment effect across demographic and baseline disease characteristics will be performed for the following subgroups:

- Gender (male vs female)
- Age (<65 vs. ≥ 65 years)
- Region (China vs rest of the world)
- Type of prior treatment received (lenvatinib vs others)
- Number of prior VEGFR targeted therapies (1 vs 2)
- Baseline TSH (≤0.5 vs >0.5 mIU per liter)
- Baseline thyroglobulin (≤10 vs. >10 ng/ml)
- ECOG ((1 and 2) vs 0)

Additional subgroup analyses with different definition or variables as well as additional supplementary analyses (e.g., handling potential impact of any COVID-19 events on treatment effect) may be conducted for PFS.

2.5.7 Supportive analyses

As a supportive analysis, PFS as per local investigator assessment will be analyzed using the same analytical conventions as the primary analysis.

As a supportive analysis, the number of patients censored and reason for censoring will be summarized by treatment group using descriptive statistics, presented separately for BIRC and local assessment.

Concordance analysis of PFS using RECIST 1.1

Cross-tabulation of 'PFS by central radiology' vs 'PFS by investigator' by PFS event type (i.e., 'death', 'PD', 'censor' for each of the two sources resulting in a 3-by-3 table) and by treatment will be constructed to investigate discordance between the two sources on a patient-by-patient basis. Discordance rate between central radiology and investigator will be calculated and presented as percentage as follows: 100 x ($n_{13} + n_{23} + n_{31} + n_{32}$) / N by treatment group. A cross-tabulation will be produced displaying the PFS timing for the local investigators' assessment compared to the BIRC assessment. For progression assessments, the frequency and percentage of subjects with complete agreement, progression later, progression earlier, and cases where progression was called by one method and censored by the other will be displayed. Similarly, if censoring was recorded, the frequency and percentage of subjects with complete agreement, censoring called later, censoring called earlier, and cases where censoring was called by one method and progression was called by the other method will be displayed.

Table 2-10 Comparison of PFS using RECIST 1.1 between investigator and BIRC

Investigator PFS	BIRC PFS result		
	Death	PD	Censor
Death	n_{11}	n_{12}	n_{13}
PD	n_{21}	n_{22}	n_{23}
Censor	n_{31}	n_{32}	n_{33}

2.5.8 Crossover

The PFS based on investigator assessment may be performed on the population of placebo patients who crossover to dabrafenib plus trametinib treatment, if there is a sufficient number of patients, and will be defined as the time from the day of the first dose of dabrafenib and trametinib treatment to the date of the first documented disease progression according to RECIST 1.1 as per investigator assessment, or death due to any cause.

Subgroup analyses may be performed for Chinese patients from sites in mainland China.

2.6 Analysis supporting key secondary objectives

2.6.1 Key secondary endpoint(s)

The key secondary objectives in this study are to compare the two treatment groups with respect to ORR and OS.

A hierarchical testing strategy will be used to control the overall type I error rate: ORR will only be formally tested and interpreted if the primary analysis of PFS is statistically significant. If the ORR achieves statistical significance, then the OS will be tested and interpreted.

Please refer to Section 2.14 for full description of the ORR and OS testing strategy.

Overall Response rate (ORR)

Overall response rate (ORR) is defined as the proportion of patients with best overall response (BOR) of confirmed complete response (CR) or partial response (PR), as per BIRC assessment and according to RECIST 1.1 (see Section 16.1 in the protocol for details). Of note, tumor assessments after the start of open-label dabrafenib and trametinib treatment following crossover for placebo patients are not considered in the BOR derivation. Responses before new anti-cancer therapy are considered.

Overall Survival (OS)

Overall survival (OS) is defined as the time from the date of randomization to the date of death due to any cause. A cut-off date will be established for each analysis of OS. All deaths occurring on or before the cut-off date in the FAS will be used in the OS analysis. If a patient is not known to have died, then OS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

2.6.2 Statistical hypothesis, model, and method of analysis

Overall Response rate (ORR)

The following null hypothesis of no difference in ORR based on BIRC assessment using RECIST 1.1 between dabrafenib plus trametinib and placebo treatment arms will be tested using Cochran-Mantel-Haenszel (CMH) test at a 1-sided significance level of 0.025, stratified by the randomization stratification factors: prior lenvatinib treatment (yes vs no) and number of prior VEGFR targeted therapies (1 vs 2).

H_{02}: $\theta_{1R} - \theta_{2R} = 0\%$ vs. H_{A2}: $\theta_{1R} - \theta_{2R} > 0\%$

where θ_{1R} and θ_{2R} are the ORR for the dabrafenib plus trametinib and placebo arms, respectively. ORR will be calculated based on the FAS, according to the ITT principle and strata assigned at randomization as per IRT.

The difference in ORR and its 95% confidence interval will be reported.

ORR and its 95% confidence interval based on the exact binomial distribution will be presented by treatment group.

Overall Survival (OS)

A hierarchical testing procedure will be adopted for OS analysis. The OS analysis will be analyzed via a two-look group sequential scheme (details are described in Section 2.14).

The following null hypothesis will be tested at one-sided 2.5% level of significance:

H_{03}: $\theta_2 \geq 1$ vs. H_{A3}: $\theta_2 < 1$

where θ_2 is the OS hazard ratio (dabrafenib plus trametinib versus placebo).

The distribution of OS will be estimated using the Kaplan-Meier method and compared between the two treatment groups using a stratified log-rank test at one-sided cumulative 2.5% level of significance, based on the FAS population. The median OS and OS Kaplan-Meier estimate at different timepoints along with 95% confidence intervals (CIs) will be presented by treatment

arm. A Cox regression model stratified by randomization stratification factors will be used to estimate the hazard ratio (HR) of OS, along with 95% CI based on the Wald test. Stratification will be based on the randomization stratification factors: prior lenvatinib treatment (yes vs no) and number of prior VEGFR targeted therapies (1 vs 2).

Recognizing potential confounding effect of crossover on OS, an attempt may be made to correct estimates with an appropriate model method, for example using Rank Preserving Structural Failure Time (RPSFT) model by Robins and Tsiatis (1991).

2.6.3 Handling of intercurrent events

The intercurrent events for the secondary estimand and handling strategies are described in Section 1.2.2.

2.6.4 Handling of missing values not related to intercurrent event

Overall Survival

If a patient is not known to have died at the time of analysis cut-off, then OS will be censored at the date of last known date patient was alive, i.e., last contact date (see Table 2-5).

2.6.5 Supportive analyses

As a supportive analysis, ORR per local investigator assessment will be analyzed using the same method as ORR by BIRC.

Disease Control Rate (DCR) defined as the proportion of patients with confirmed BOR of CR or PR or SD will be reported.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.7 Analysis supporting secondary objectives

The other secondary objective is to evaluate DOR of dabrafenib plus trametinib versus placebo.

2.7.1 Secondary endpoint(s)

Duration of response (DOR)

Duration of response only applies to patients whose best overall response is confirmed complete response (CR) or partial response (PR) according to RECIST 1.1 based on BIRC assessment. The start date is the date of first documented confirmed response of CR or PR, and the end date is defined as the date of the first documented confirmed progression or death due to any cause. Patients continuing without progression or death due to any cause will be censored at the date of their last adequate tumor assessment before the start of new antineoplastic therapy, if any.

2.7.2 Statistical hypothesis, model, and method of analysis

Duration of response (DOR)

DOR will be listed and summarized by treatment group for all patients in the FAS with confirmed BOR of CR or PR. The distribution of duration of response will be estimated using

the Kaplan-Meier method and the median duration of response will be presented along with 95% confidence interval by treatment groups. No inferential analysis that compares duration of response between the two treatment groups will be performed.

2.7.3 Handling of missing values not related to intercurrent event

Duration of response (DOR)

The date of last adequate tumor assessment is the date of the last tumor assessment with overall response of CR, PR or SD before an event or a censoring reason occurred. In this case the last tumor evaluation date at that assessment will be used. If no post-baseline assessments are available (before an event or a censoring reason occurred) then the date of randomization/start date of treatment will be used.

In particular, DOR will be censored at the last adequate tumor assessment if one of the following occurs: absence of event; the event occurred after two or more missing tumor assessments; before the start of the new anti-cancer therapy. The term “missing adequate tumor assessment” is defined as a tumor assessment (TA) not performed or tumor assessment with overall response of “UNK”. The rule to determine number of missing TAs is based on the time interval between the date of last adequate tumor assessment and the date of an event. If the interval is greater than twice the protocol-specified interval between the TAs and 2 times the protocol-allowed time window around assessments, then the number of missing assessments will be 2 or more.

Refer to Table 2-9 for censoring and event date options and outcomes for DOR and PFS analysis.

2.7.4 Supplementary analyses

As supportive analyses, DOR will be performed as per local investigator review using the same analytical approach than for the analyses based on BIRC assessment.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.8 Safety analyses

For all safety analyses except ocular event analyses, the Safety Set will be used. For the ocular event analyses, both the Safety Set and the Ocular Event Evaluable Set will be used. All listings and tables will be presented by treatment group.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). In addition, a separate summary for death including on-treatment and post treatment deaths will be provided. In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period (treatment-emergent AEs). Selected key analyses of AEs will also be summarized based on the Crossover Population Set.

The overall observation period will be divided into three mutually exclusive segments:

1. Pre-treatment period: from day of patient's informed consent to the day before first administration of the study treatment

2. On-treatment period: from day of first administration of the study treatment to the earlier of
   - 30 days after date of last administration of the study treatment
   - day prior to the start of open-label dabrafenib and trametinib treatment following crossover from placebo
3. Post-treatment period: starting at day 31 after last administration of study treatment or at the first day of open-label dabrafenib and trametinib treatment.

All AEs, deaths, and serious adverse events (including those from the pre and post-treatment periods) will be listed and those collected during the pre-treatment and post-treatment period will be flagged

2.8.1 Adverse events (AEs)

AE summaries will include all AEs occurring during the on-treatment period. All AEs collected in the AE (e)CRF page will be listed along with the information collected on those AEs e.g. AE relationship to study drug, AE outcome etc. AEs with start date outside of on-treatment period will be flagged in the listings.

AEs will be summarized by number and percentage of patients having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding. A patient with multiple occurrences of an AE will be counted only once in the respective AE category. A patient with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AEs with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

In AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on their frequency in the investigational group (dabrafenib and trametinib).

The following adverse event summaries will be produced by treatment group: overview of adverse events and deaths (number and % of patients with any AEs, treatment-related AEs, SAEs, fatal AEs, AEs leading to discontinuation, AEs leading to dose reduction/interruption, or AEs requiring additional therapy), AEs by SOC and PT, summarized by relationship (all AEs and AEs related to study treatment), seriousness (SAEs and non-SAEs), leading to treatment discontinuation, leading to dose adjustment and/or interruption, leading to dose reduction (for dabrafenib and/or trametinib only), requiring additional therapy, and leading to fatal outcome. In addition, a summary of serious adverse events with number of occurrences will be produced (an occurrence is defined as >1 day between start and prior end date of record of same preferred term).

Subgroup analyses of AEs will be performed for Chinese patients from sites in mainland China.

For legal requirements of clinicaltrials.gov and EudraCT, two required tables for on-treatment adverse events which are not SAEs with an incidence greater than and equal to 5% and on-treatment deaths and SAEs suspected to be related to study treatment will be provided by system organ class and preferred term in the Safety Set.

2.8.1.1 Adverse events of special interest / grouping of AEs

All AE groupings for a clinical program are stored in the Compound electronic Case Retrieval Strategy sheet (eCRS) with clear versioning and reference to the MedDRA version used.

All AESI definitions or AE groupings need to be specified in the CRS. If a CRS update is necessary, the final version needs to be available in a reasonable time ahead of the DBL. The CRS version should be included in a footnote of the AESI tables.

Data analysis of AESIs

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to the combination of dabrafenib and trametinib. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad.

For each specified AESI, number and percentage of patients with at least one event of the AESI occurring during on-treatment period will be summarized.

Summaries of these AESIs will be provided by treatment group (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose adjustment/interruption, hospitalization, death, requiring medication etc.).

A listing of all grouping levels down to the MedDRA preferred terms used to define each AESI will be generated.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.8.2 Deaths

All deaths (on-treatment and post-treatment) will be summarized overall and separately by treatment group, system organ class and preferred term.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

If study indication is the primary reason for death (and not coded accordingly in the database), it must be included in the summary table. All deaths will be listed; post treatment deaths will be flagged. The death summaries cover patients from the Safety Set. A separate listing of deaths prior to starting treatment will be provided for all screened subjects.

2.8.3 Laboratory data

All laboratory data will be listed by treatment group, patient, and visit and if normal ranges are available abnormalities will be flagged. Summary statistics will be provided by treatment group. Shift tables using the low/normal/high/(low and high) classification will be used to compare baseline to the worst on-treatment value.

Grading of laboratory values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE

grades will be based on the observed laboratory values only, clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher. Grade 5 (death) is not applicable.

For laboratory tests where grades are not defined by CTCAE v4.03, results will be categorized as low/normal/high based on laboratory normal ranges.

Laboratory data from all sources (central and local laboratories) will be combined. The summaries will include all assessments available for the lab parameter collected no later than 30 days after the last study treatment administration date (see Section 2.1.1). All laboratory assessments will be listed and those collected later than 30 days after the last study treatment/exposure date will be flagged in the listings.

The following listings/summaries will be generated separately for hematology, and biochemistry tests:

- Listing of all laboratory data with values flagged to show the corresponding CTCAE v4.03 grades if applicable and the classifications relative to the laboratory normal ranges

For laboratory tests where grades are defined by CTCAE v4.03:

- Worst post-baseline CTCAE grade (regardless of the baseline status). Each patient will be counted only once for the worst grade observed post-baseline.
- Shift tables using CTCAE v4.03 grades to compare baseline to the worst on-treatment value.

For laboratory tests where grades are not defined by CTCAE v4.03:

- Shift tables using the low/normal/high/ (low and high) classification to compare baseline to the worst on-treatment value.
- Trends of lab parameter values over time (baseline and selected on-treatment timepoints) should be displayed via boxplots based on time windows and corresponding tables displaying the statistics used for the box plots by the selected time points.

The following listings will be produced for the laboratory data:

- Listings of all laboratory data, with CTCAE grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.
- Listing of all CTCAE grade 3 or 4 laboratory toxicities

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

Liver parameters

Liver parameters of interest are total bilirubin (TBL), ALT, AST and alkaline phosphatase (ALP). The number (%) of patients with worst post-baseline values as per Novartis Liver Toxicity guidelines will be summarized:

The following summaries will be produced:

- ALT or AST > 3xULN

- ALT or AST > 5xULN
- ALT or AST > 8xULN
- ALT or AST > 10xULN
- ALT or AST > 20xULN
- TBL > 2xULN
- TBL > 3xULN
- ALT or AST > 3xULN & TBL > 2xULN
- ALT or AST > 3xULN & TBL > 2xULN & ALP < 2xULN (potential Hy's law)
- ALT or AST > 3xULN & TBL > 2xULN & ALP >= 2xULN

For patients with AST/ALT and bilirubin within normal levels at baseline, potential Hy's Law cases were defined as those patients with occurrence of AST or ALT >3xULN and TBL >2xULN, and ALP <2xULN at initial presentation during the on-treatment period. The criteria relating to combined elevations of AST (or ALT) and TBL were based on the peak values at any post-baseline time for a patient.

For patients with abnormal ALT or AST baseline values, a clinically significant liver safety signal corresponding to Hy's law was defined by: [ALT or AST >3*baseline] OR [ALT or AST >8*ULN], whichever was lower, combined with [TBIL >2*baseline AND >2*ULN] and ALP < 2xULN.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

A standard 12-lead ECG (single) will be performed according to the relevant Visit Evaluation Schedule. Electrocardiograms (ECGs) must be recorded after 10 minutes rest in the supine position to ensure a stable baseline. The preferred sequence of cardiovascular data collection during study visits is ECG collection first, followed by vital signs, and blood sampling. Additional, unscheduled, ECGs may be performed at the discretion of the investigator at any time during the study as clinically indicated. Unscheduled ECGs with clinically significant findings should be collected in triplicate.

ECHO or MUGA (ECHO is preferred) will be performed to assess cardiac ejection fraction in regular intervals according to the relevant schedule. The same procedure (ECHO or MUGA) should be performed at screening/cross over baseline and at follow-up visit(s).

Data handling

In case the study requires ECG replicates at any assessment, the average of the ECG parameters at that assessment should be used in the analyses.

Data analysis

12-lead ECG including HR, PR, QRS, QT, QTcF, and RR intervals will be obtained for each subject during the study.

The number and percentage of subjects with notable ECG values will be presented by treatment arm.

- QT, QTcF
  - New value of > 450 and ≤ 480 ms
  - New value of > 480 and ≤ 500 ms
  - New value of > 500 ms
  - Increase from baseline of > 30 ms to ≤ 60 ms
  - Increase from baseline of > 60 ms
- HR
  - Increase from baseline >25% and to a value > 100 bpm
  - Decrease from baseline >25% and to a value < 50 bpm
- PR
  - Increase from baseline >25% and to a value > 200 ms
  - New value of > 200 ms
- QRS
  - Increase from baseline >25% and to a value > 120 ms
  - New values of QRS > 120 ms

For each of the ECG parameters (QT, QTc, QRS, HR and PR intervals), descriptive statistics at baseline, at each post-baseline time point and changes from baseline at each post-baseline time point will be summarized.

A listing of all ECG assessments will be produced by treatment arm and notable values will be flagged. A separate listing of only the subjects with notable ECG values may also be produced. In the listing, the assessments collected during the post-treatment period will be flagged.

Unscheduled ECG measurements will not be used in computing the descriptive statistics for change from baseline at each post-baseline time point. However, they will be used in the analysis of notable ECG values.

For left ventricular ejection fraction (LVEF), descriptive statistics at baseline, at each post-baseline time point and changes from baseline at each post-baseline time point will be summarized by treatment group. The number and percentage of subjects with LVEF ≥ 10% from Baseline and below the institutional lower limit of normal (LLN) will be presented by treatment arm.

Subgroup analyses will be performed for ECG and LVEF for Chinese patients from sites in mainland China.

2.8.4.2 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: height (cm), weight (kg), body temperature (°C), pulse rate (beats per minute), systolic and diastolic blood pressure (mmHg).

Data handling

Vital signs collected on treatment will be summarized. Values measured outside of on-treatment period will be flagged in the listings.

Data analysis

For analysis of vital signs the clinically notable vital sign criteria are provided in Table 2-11 below.

Table 2-11 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	Above normal value	**Below normal value**
Weight (kg)	Increase >10% from baseline	Decrease >10% from baseline
Systolic blood pressure (mmHg)	≥180 with increase from baseline of ≥20	≤90 with decrease from baseline of ≥20
Diastolic blood pressure (mmHg)	≥105 with increase from baseline of ≥15	≤50 with decrease from baseline of ≥15
Pulse rate (bpm)	≥100 with increase from baseline of >25%	≤50 with decrease from baseline of >25%
Body temperature (°C)	≥39.1	-

The number and percentage of subjects with notable vital sign values (high/low) will be presented by treatment arm. Descriptive statistics will be tabulated for baseline, at each post-baseline time point and changes from baseline at each post-baseline time point for each vital sign measure.

A listing of all vital sign assessments will be produced by treatment arm and notable values will be flagged. A separate listing of only the subjects with notable vital sign values may also be produced. In the listing, the assessments collected outside of on-treatment period will be flagged.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.8.4.3 ECOG PS

ECOG PS will be summarized with shift tables showing change from baseline to the worst observed post-baseline value.

Subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.8.4.4 Ophthalmic examinations

Listings of all ophthalmic examinations will be presented by treatment group. In the listings, the assessments collected outside of on-treatment period will be flagged.

2.8.5 Serous retinopathy ocular events

The study will be used to estimate the incidence of, and to characterize trametinib-associated serous retinopathy ocular events, to meet a Post Marketing Requirement (PMR).

The intercurrent events for the estimand and handling strategies are described in Section 1.2.2.

Both Safety Set and Ocular Event Evaluable Set will be used.

As a secondary estimand and supplementary of secondary estimand, incidence, type and severity of ocular events using serous retinopathy grouping term will be summarized by treatment arm using the Ocular Event Evaluable Set and the Safety Set. The 95% confidence interval of the incidence of serous retinopathy event will be as well presented. The following AEs related to serous retinopathy ocular events will be presented by treatment group: overview of AEs (number and %, treatment-related AE, SAE, AE leading to discontinuation, AE leading to dose reduction/interruption, or AE requiring additional therapy), AEs by SOC and PT, summarized by relationship (AEs related to study treatment only), seriousness (SAEs and non-SAEs), grade, leading to treatment discontinuation, leading to dose adjustment and/or interruption, leading to dose reduction, requiring additional therapy. Serous retinopathy will be defined based on a specific case retrieval strategy for ocular events as serous retinopathy ocular events, listing the appropriate preferred terms falling into the definition, which are listed as follows:

- Chorioretinopathy
- Detachment of macular retinal pigment epithelium
- Detachment of retinal pigment epithelium
- Macular detachment
- Maculopathy
- Metamorphopsia
- Retinal detachment
- Retinal disorder
- Retinal pigment epitheliopathy
- Retinopathy
- Subretinal fluid
- Serous retinopathy
- Central serous chorioretinopathy
- Serous retinal detachment
- Retinal pigment epithelial tear
- Retinal microangiopathy
- Proliferative vitreoretinopathy
- Retinal pigment epithelium change

Above PT list shows the current list and reflects PT consolidations for updated MedDRA version 27.1. Should there be further MedDRA updates, Novartis intends to use the most updated MedDRA version and include the PTs that encompass the current serous retinopathy PT list.

Listings of visual acuity, tonometry, visual field, fundoscopy, slit lamp examination, and optical coherence tomography will be provided for the Ocular Event Evaluable Set.

Listings of central read optical coherence tomography data will also be provided.

For the purpose of the PMR, an output of serous retinopathy incidence based on the pooled blinded clinical adverse event data will be provided 1 year after FPFV, and annually thereafter until the final report submission. The serous retinopathy events for this annual PMR update will also be based on the preferred terms grouping listed above.

A subgroup analyses will be performed for Chinese patients from sites in mainland China.

2.8.6 Crossover

Selected AE summaries listed above for the Safety Set will be also provided using the Crossover Population Set. All crossover patients will be used until end of treatment.

[redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted]

[redacted]

[redacted]

[redacted] [redacted]

[redacted]

[redacted] [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

2.14 Interim analysis

Primary endpoint: Progression free survival (PFS)

There is no interim analysis (IA) for PFS. A minimum of 77 events need to be observed for 90% power, however, sample size needed to be increased to fulfill the ocular-event analysis regulatory requirement. Therefore, the primary analysis will be performed after approximately 95 PFS events have been reported as per BIRC assessment, and all randomized patients should have approximately 16 weeks of follow-up or have discontinued before. Formal testing of the primary endpoint with full alpha will be performed at the primary analysis.

Key secondary endpoints : Overall Response Rate (ORR) and Overall Survival (OS)

A hierarchical testing procedure will be adopted and the statistical test for ORR will be performed only if the primary analysis of PFS is statistically significant. OS will be tested only if the primary analysis of PFS and key secondary ORR endpoint are statistically significant.

For ORR, there is no interim analysis planned. The analysis will be performed at the time of the primary analysis for PFS (and provided PFS is significant).

For OS, a maximum of two analyses are planned:

- At the time of the primary analysis for PFS (and provided PFS and ORR are significant), at which point a total of approximately 39 deaths are expected (around 31 months from first patient randomized)
- At the time of final analysis for OS when all randomized patients will have been followed for at least 3 years.

The type I error rate for OS testing will be controlled by using a one-sided 2-look group sequential design. Specifically, an α-spending function according to Lan-DeMets (Pocock) as implemented in East (6.4) Lan and DeMets DL 1983 to spend sufficient alpha and increase the chance to be positive at interim (given later confounding of OS analysis due to crossover), along with the testing strategy outlined above will be used to maintain the overall type I error probability. This guarantees the protection of the 2.5% overall level of significance across the two hypotheses and the repeated testing of the OS hypotheses at the interim and the final analysis (Glimm et al 2010). The final OS analysis will be performed once all patients will have been followed up for at least 3 years, at which point, approximately 82 deaths are expected to have occurred. This number was used for the alpha-spending calculations and determining information fraction used at the interim analysis.

The trial allows for the stopping of the study for a superior OS result, provided the primary endpoint PFS and key secondary endpoint ORR have already been shown to be statistically significant favoring the test treatment arm. Further, the exact nominal p-values that will need to be observed to declare statistical significance at the time of these analyses for OS will depend on the number of OS events that have been observed at the time of these analyses and the α for OS already spent at the time of the earlier interim analysis.

The projected timing of interim and final OS analysis is summarized in Table 2-13.

At the time of primary analysis, PFS, ORR and interim OS analysis will be performed by the Sponsor's clinical team. Investigators and patients will remain blinded to study treatment as much as possible and if patients do not crossover to dabrafenib and trametinib arm. All patients will continue to be followed for OS until the final OS analysis.

Table 2-13 Estimated timelines for interim and final OS analyses

Months after randomization of the first patient (approximation)	# PFS Events	Cumulative power against a PFS hazard ratio of 0.455	# OS events	Cumulative[b] power against OS hazard ratio of 0.7
31	95 (100%)	96.2%	39 (47.6%) [a]	15.1%
62	-	-	82 (100%) [a]	31.1%

a: Interim/final analysis for OS will only be performed if the final analysis for PFS and key secondary ORR analysis are significant

b: Power conditional on PFS and ORR being significant

NOTE: Simulation is performed in East (6.4) with number of simulations = 10,000 and randomization seed =1234, under alternative hypothesis.

3 Sample size calculation

3.1 Primary endpoint(s)

The sample size calculation is based on the question of interest related to the primary estimand. The hypotheses to be tested and details of the testing strategy are described in Section 2.5.2.

Based on Lenvatinib (Schlumberger et al 2015) and Sorafenib data (Brose et al 2014), including mostly naïve patients (without receiving any prior treatment regimen), the median PFS in the control arm is expected to be around 5 months for this study population.

Under the assumption that the median PFS in the control arm is 5 months, it is expected that treatment with dabrafenib and trametinib will result in a 55% reduction in the PFS hazard rate (corresponding to an increase in median PFS from 5 months to 11 months under the exponential model assumption). If the true HR=0.455 (under the alternative hypothesis), the expected number of 95 PFS events at the time of the analysis (based on a minimum 16-week follow-up time for all patients) will have 95% power at a one-sided 2.5% level of significance to reject the null hypothesis (HR=1) using a log-rank test. Considering a recruitment period of approximately 26 months from the start of the study with an accrual rate of 1 patient/month for first 4 months, 3 patients/month from 4 to 8 months, 6 patients/month for 9 to 11 months and 8 patients/month thereafter, along with an assumed 8% dropout rate/year, approximately 150 patients will need to be randomized to the two treatment arms in a 2:1 ratio. Given the above assumptions, it is estimated that the cutoff for the data analysis will occur at approximately 31

months from the date of the first patient randomized in the study at which point approximately 95 events are expected. Of note, to achieve 90% power only 77 PFS events would have been needed under the same assumptions. However, sample size needed to be increased to fulfill the ocular-event analysis regulatory requirement. This results in an increase of the power to ensure also that all patients have been randomized and followed for a minimum 16-week follow-up time, at time of primary analysis. The sample size calculation was conducted with software package East 6.4.

3.2 Secondary endpoint(s)

Serous retinopathy ocular events incidence

In order to fulfill regulatory requirement, patients will be followed for serous retinopathy ocular events assessment (a set of AEs previously reported in patients taking trametinib). No comparison with placebo arm will be done, only the incidence of serous retinopathy events will be provided for the two treatment arms, with specific interest in the dabrafenib and trametinib arm, as part of a secondary endpoint. The sample size of 150 total randomized patients will provide acceptable precision for these serous retinopathy ocular events rates. Approximately 100 patients would be randomized in the dabrafenib and trametinib arm to get ~90 evaluable patients in dabrafenib and trametinib arm assuming a dropout rate of ~10%. With 90 evaluable patients, if the true serous retinopathy ocular events rate with dabrafenib and trametinib treatment is 10% or less, there will be at least 71.3% probability to show that the upper bound of the 95% CI is below 20% (i.e., the current serous retinopathy ocular events rate reported for other MEK-inhibitors treatments). Table 3-1 below shows various scenarios for serous retinopathy ocular event rates (and corresponding number of observed events), exact Clopper-Pearson 95% confidence intervals among 90 and 150 evaluable patients. These calculations were made using R3.4.3.

Table 3-1 95% confidence intervals for serous retinopathy ocular events rates among 90 and 150 evaluable patients

# of patients evaluable in dabrafenib and trametinib arm	# of patients with events	Event rate (%)	95% CI (%)
90	4	4.4	(1.2 - 11.0)
	7	7.8	(3.2 - 15.4)
	9	10.0	(4.7 - 18.1)
	10	11.1	(5.5 - 19.5)
	11	12.2	(6.3 - 20.8)
	14	15.6	(8.8 - 24.7)
	18	20.0	(12.3 - 29.8)
150	5	3.3	(1.1 - 7.6)
	10	6.7	(3.2 - 11.9)
	15	10.0	(5.7- 16.0)
	20	13.3	(8.3 - 19.8)
	30	20.0	(13.9 - 27.3)

Overall response rate

Overall response rate, as one the key efficacy secondary objectives, will be formally statistically tested, provided that the primary analysis of PFS is statistically significant. Based on a phase II study (Shah et al 2017), it is hypothesized that dabrafenib and trametinib would result in an overall response rate of 35%. Overall response rate in the placebo arm is expected to be around 1%. The number of patients planned to be enrolled (i.e., 150 patients), will provide sufficient power (>99%) to reject the null hypothesis of no difference in overall response rate between treatment arms, using a Cochran-Mantel-Haenszel (CMH) test at a 1-sided significance level of 0.025. The overall response rate analysis will be performed at the time of the primary analysis. These calculations were made using the software package East 6.4.

Overall survival

Overall survival, as one of the key secondary objectives, will be formally statistically tested, provided that the primary analysis of PFS and key secondary ORR endpoint are statistically significant and using a 2-look group sequential design with Lan-DeMets (Pocock) alpha spending function. Based on data (Schlumberger et al 2015, Cabanillas et al 2017), the median OS in the control arm is expected to be around 24 months. It is hypothesized that treatment with dabrafenib and trametinib will result in a 30% reduction in the hazard rate for OS, i.e., an expected hazard ratio of 0.70 (which corresponds to an increase in median OS to 34.3 months under the exponential model assumption). This study is not designed to be powered for overall survival and considering the confounding of OS due to crossover of placebo patients to the dabrafenib and trametinib arm at time of disease progression (confirmed by independent review), there will be only a low chance (29%) to reject the null hypothesis of OS hazard ratio = 1. Based on the number of patients planned to be enrolled to provide sufficient power for the primary estimand (i.e., 150 patients), assuming 10% drop-out rate per treatment arm per year, and to align with the primary PFS analysis timing after around 31 months from first patient randomized, it is estimated that approximately 39 OS events will be observed. The final analysis will be driven by calendar time and will be performed once all patients have been followed for at least 3 years. These calculations were made using the software package East 6.4.

4 Change to protocol specified analyses

Not applicable.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

The following rule should be used for the imputation of the dose end date for a given study treatment component:

Scenario 1: If the dose end date is completely missing and there is no EOT page and no death date, the patient is considered as on-going:

The patient should be treated as on-going and the cut-off date should be used as the dose end date.

Scenario 1 should not be applicable for final CSR. All patients should have EOT page complete before the Database lock for final CSR.

Scenario 2: If the dose end date is completely or partially missing and the EOT page is available:

Case 1: The dose end date is completely missing, and the EOT completion date is complete, then this latter date should be used.

Case 2: Only Year (yyyy) of the dose end date is available and yyyy < the year of EOT date:

- **Use Dec31yyyy**

Case 3: Only Year (yyyy) of the dose end date is available and yyyy = the year of EOT date:

- **Use EOT date**

Case 4: Both Year (yyyy) and Month (mm) are available for dose end date, and yyyy = the year of EOT date and mm < the month of EOT date:

- **Use last day of the Month (mm)**

Case 5: Both Year (yyyy) and Month (mm) are available for dose end date, and yyyy = the year of EOT date and mm = the month of EOT date:

- **Use EOT date**

All other cases should be considered as a data issue and the statistician should contact the data manager of the study. If imputation is needed, date should be imputed in the most conservative way. For example, if a dose interruption or reduction start date and end date are partial, then the imputation of the dates should be based on the longest possible dose interruption/reduction.

After imputation, compare the imputed date with start date of treatment, if the imputed date is ≤ start date of treatment:

- **Use the treatment start date**

Patients with missing start dates are to be considered missing for all study treatment component related calculations and no imputation will be made. If start date is missing then end-date should not be imputed.

5.1.2 AE, ConMeds and safety assessment date imputation

Table 5-1 Imputation of start dates (AE, CM, prior ANP, medical procedures, etc.) and assessments (LB, EG, VS)

Missing Element	Rule
day, month, and year	No imputation will be done for completely missing dates

day, month	• If available year = year of study treatment start date then
	  • If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY
	  • Else set start date = study treatment start date.
	• If available year > year of study treatment start date then 01JanYYYY
	• If available year < year of study treatment start date then 01JulYYYY

day	• If available month and year = month and year of study treatment start date then
	  • If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY.
	  • Else set start date = study treatment start date.
	• If available month and year > month and year of study treatment start date then 01MONYYYY
	• If available month and year < month and year of study treatment start date then 15MONYYYY

Table 5-2 Imputation of end dates (AE, CM)

Missing Element	**Rule** (*=last treatment date plus 30 days not > (death date, cut-off date, withdrawal of consent date))
day, month, and year	Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period*
day	If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

Any AEs and ConMeds with partial/missing dates will be displayed as such in the data listings.

Any AEs and ConMeds which are continuing as per data cut-off will be shown as 'ongoing' rather than the end date provided.

The above imputations are only used for analyses of time to and duration of AEs and concomitant medications.

5.1.2.1 Other imputations

Incomplete date of initial diagnosis of cancer and date of most recent recurrence

Missing day is defaulted to the 15th of the month and missing month and day is defaulted to 01-Jan.

Incomplete assessment dates for tumor assessment

All investigation dates (e.g., MRI scan, CT scan) must be completed with day, month and year. If one or more assessment dates are incomplete but other investigation dates are available,

this/these incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the latest of all investigation dates (e.g., MRI scan, CT scan) if the overall response at that assessment is CR/PR/SD/UNK. Otherwise – if overall response is progression – the assessment date is calculated as the earliest date of all investigation dates at that evaluation number. If all measurement dates have no day recorded, the 1st of the month is used. If the month is not completed, for any of the investigations, the respective assessment will be considered to be at the date which is exactly between previous and following assessment. If a previous and following assessment is not available, this assessment will not be used for any calculation.

Applying the cut-off to tumor assessment

For tumor related assessments, if an evaluation has some assessments done prior to cut-off date and others after the cut-off date, then the evaluation is considered post-cut-off date and will be excluded from analysis.

5.2 Adverse events coding/grading

AEs are coded using the MedDRA terminology.

Note: The latest available MedDRA version at the time of the analyses should be used. The MedDRA version used should be specified in the footnote of relevant tables.

AEs will be assessed according to the CTCAE v4.03.

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death. This grading system inherently places a value on the importance of an event; although there is not necessarily proportionality among grades (a grade 2 is not necessarily twice as bad as a grade 1).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI CTCAE v4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTCAE grades are given in Novartis internal criteria for CTCAE grading of laboratory parameters. The latest available version of the document based on the underlying CTCAE v4.03 at the time of analysis will be used.

For laboratory tests where grades are not defined by CTCAE v4.03, results will be graded by the low/normal/high classifications based on laboratory normal ranges.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. Grade 5 will not be used. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

Imputation Rules

CTC grading for blood differentials is based on absolute values. However, this data may not be reported as absolute counts but rather as percentage of WBC.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

The following rules will be applied to derive the WBC differential counts when only percentages are available for an xxx differential

$$xxx\ count = (WBC\ count) * \left(\frac{xxx\ \%\ value}{100}\right)$$

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

$$Corrected\ Calcium\ \left(\frac{mg}{dL}\right) = Calcium\left(\frac{mg}{dl}\right) - 0.8\left[Albumin\ \left(\frac{g}{dL}\right) - 4\right]$$

In order to apply the above formula, albumin values in g/L will be converted to g/dL by multiplying by 0.1, calcium values in mmol/L will be converted to mg/dL by dividing by 0.2495. For calculation of laboratory CTC grades 0 and 1, the normal range for derived corrected calcium is set to the same limits (in mg/dL) as for calcium.

CTC grades for the derived absolute WBC differential counts (neutrophils, lymphocytes) and corrected calcium will be assigned as described above for grading.

5.4 Statistical models

5.4.1 Analysis supporting primary objective(s)

Analysis of time to events Data

Hypothesis and test statistic

The null hypothesis stating that PFS survival distributions of the two treatment groups are equivalent will be tested against the one-sided alternative.

The following statistical hypotheses will be tested:

$$H_{01}: \theta_1 \geq 1 \quad \text{vs.} \quad H_{A1}: \theta_1 < 1$$

where θ_1 is the PFS hazard ratio (dabrafenib and trametinib vs placebo).

Stratified log-rank test adjusting for the strata used in the randomization will be implemented as follows: In each of the K strata separately, the LIFETEST procedure with STRATA statement including only the treatment group variable and with the TIME statement will be used to obtain the rank statistic S_k and variance var(S_k) where k=1, 2, …, K. The final test statistics will then be reconstructed as follows:

$Z=[S_1 +\ldots+ S_K]/\sqrt{[var(S_1) +\ldots+ var(S_k)]}$.

The one-sided p-value will be obtained using a Z statistic.

Kaplan-Meier estimates

An estimate of the survival function in each treatment group will be constructed using Kaplan-Meier (product-limit) method as implemented in PROC LIFETEST with METHOD=KM option. The PROC LIFETEST statement will use the option CONFTYPE=LOGLOG.

Median survival for each treatment group will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the method of (Brookmeyer and Crowley 1982). Kaplan-Meier estimates of the survival function with 95% confidence intervals at specific time points will be summarized. The standard error of the Kaplan-Meier estimate will be calculated using Greenwood's formula (Collett 1994).

Hazard ratio

Hazard ratio will be estimated by fitting the Cox proportional hazards model using SAS procedure PHREG (with TIES=EXACT option in the MODEL statement).

A stratified unadjusted Cox model will be, i.e., the MODEL statement will include the treatment group variable as the only covariate and the STRATA statement will include stratification variable(s).

Hazard ratio with two-sided 95% confidence interval will be based on Wald test.

Note: Ideally, the hazard ratio and the confidence interval should be derived by a method consistent with the p-value calculation, i.e., in this case with log-rank test. This requirement would lead to the score test based intervals. However, since these intervals are not available in the SAS procedure PHREG, Wald test based intervals will be used.

Treatment of ties

Note: Ideally, the ties handling method used in LIFETEST and PHREG procedures should be consistent. However, since the main purpose of employing the PHREG procedure is to produce a hazard ratio with confidence interval and this cannot be obtained in a way consistent with log-rank based p-value produced by LIFETEST, it is recommended that the PHREG procedure should use a ties handling method which is considered optimal in given setting regardless of the consistency between LIFETEST and PHREG procedures.

The STRATA statement in LIFETEST procedure will be used to analyze time to event data with ties. The PHREG procedure in SAS with option TIES = EXACT will be used to fit the Cox proportional hazards model.

Checking proportionality of hazard assumption

Plots (SURVIVAL LOGSURV LOGLOGS) generated by LIFETEST procedure in SAS will be used to provide visual checks of the proportional hazard assumption:

- SURVIVAL plots estimated survivor functions. The shape of the curves should be basically the same if hazards are proportional.
- LOGSURV plots the cumulative hazard functions. The larger cumulative hazard should be a multiple of smaller if hazards are proportional.

- LOGLOGS plots log (cumulative hazard). The LOGLOG plot will show parallel curves if hazards are proportional.

[redacted]

5.4.2 Analysis supporting secondary objective(s)

For OS and DOR, the same instructions described in Section 5.4.1 apply.

The null hypothesis of equality of response rate in the two treatment arm will be tested against one-sided alternative. The statistical hypotheses are:

H_{02}: $\theta_{1R} - \theta_{2R} = 0\%$ vs. H_{A2}: $\theta_{1R} - \theta_{2R} > 0\%$ where θ_{1R} and θ_{2R} are the ORR for the dabrafenib plus trametinib and placebo arms, respectively.

The Mantel-Haenszel chi-square test X^2_{MH} (implemented via SAS procedure FREQ with CMH option in the TABLES statement) will be used to test the difference in response rates between the treatment arms at one-sided 2.5% level of significance, stratified by the randomization stratification factors. The p-value corresponding to the CMH test for "general association" will be used which follows a Chisquare distribution with one degree of freedom.

If the sampling assumptions for chi-square test is not met, the exact Cochran-Mantel-Haenszel test (implemented via SAS procedure MULTTEST) will be used. The test is performed by running a stratified version of the Cochran-Armitage permutation test. In studies with stratified randomization, the chi-square approximation is considered appropriate for the X^2_{MH} statistics if the rule of Mantel and Fleiss (1980) is satisfied.

Confidence interval for response rate

Responses will be summarized in terms of percentage rates with $100(1-\alpha)\%$ confidence interval using exact binomial confidence interval (implemented using SAS procedure FREQ with EXACT statement for one-way table (Clopper and Pearson 1934).

6 Reference

1. Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41.
2. Brose MS, Nutting CM, Jarzab, et al (2014) Sorafenib in locally advanced or metastatic, radioactive iodine-refractory, differentiated thyroid cancer: a randomized, double-blind, phase 3 trial. Lancet. 384: 319–328.
3. Cabanillas ME, de Souza JA, Geyer S, et al (2017) Cabozantinib As Salvage Therapy for Patients With Tyrosine Kinase Inhibitor-Refractory Differentiated Thyroid Cancer: Results of a Multicenter Phase II International Thyroid Oncology Group Trial. J. Clin. Oncol. 17: 3315-3321.
4. CHMP guideline on adjustments for baseline covariates in clinical trials (2015)
   http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500184923.pdf
5. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26, 404-413.
6. Collett D (1994). Modelling survival data in medical research. London, Chapman & Hall.
7. Glimm E, Maurer W and Bretz F (2010). Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine, 29, 219-228.
8. ICH E9(R1) Harmonized Guideline: addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Final version on 20 November 2019.
9. Lan KKG, DeMets DL (1983). Discrete sequential boundaries for clinical trials. Biometrika, 70, 649-53.
10. Mantel N, Fleiss JL (1980). Minimum expected cell size requirements for the Mantel-Haenszel one-degree-of-freedom chi-square test and a related rapid procedure. American Journal of Epidemiology, 112:129-134.
11. Robins JM and Finkelstein DM (2000). Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics, 56(3),779-88.
12. Robins JM and Tsiatis AA (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models (RPSFTM). Commun. Statist.-Theory Meth., 20(8), 2609-31.
13. Schlumberger M, Tahara M, Wirth LJ, et al (2015) Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. N. Engl. J. Med. 372(7): 621-30.
14. Shah MH, Wei L, Wirth LJ, et al (2017) Results of randomized phase II trial of dabrafenib versus dabrafenib plus trametinib in BRAF-mutated papillary thyroid carcinoma. Journal of Clinicasl Oncology. Vol 35: Issue 15 Suppl.