European Thoracic Oncology Platform | c/o IBCSG | Effingerstrasse 40 | 3008 Bern | Switzerland | www.etop.ch

etop
Information | Research

ETOP 10-16 BOOSTER

A randomised phase II trial of osimertinib and bevacizumab versus osimertinib alone as second-line treatment in stage IIIb-IVb NSCLC with confirmed EGFRm and T790M

Statistical Analysis Plan (SAP)
Final efficacy analysis

A clinical trial of ETOP

SAP VERSION	2.1
SAP DATE	23.12.2020

Contents

INTRODUCTION 4
1 Trial oversight 5
1.1 Objectives 5
1.2 Endpoints 6
1.3 Most important eligibility criteria 6
1.4 Trial treatment 7
1.5 Statistical design, sample size & power 7
1.6 Total trial duration 8
2 Statistical considerations for final analysis 9
2.1 Analysis timing 9
2.2 Study's endpoints 9
2.2.1 Primary endpoint 9
2.2.2 Secondary endpoints 9
2.2.3 Exploratory endpoints 9
2.2.4 Correlative endpoints 10
2.3 (Serious) Adverse Events 11
2.4 Analysis populations 12
3 Primary efficacy analysis of progression-free survival (PFS) 14
4 Additional secondary analysis 16
4.1 Patient accrual, balance of stratification factors and baseline characteristics 16
4.2 Follow-up information and treatment administration 16
4.3 Secondary (efficacy) and exploratory endpoints 17
4.3.1 Overall survival 17
4.3.2 Objective response rate & disease control rate 18
4.3.3 Duration of response and duration of clinical benefit 18
4.3.4 Subgroup analysis 18
4.4 Sensitivity efficacy analysis 19
4.5 Safety analysis 19

5 Technical details 21
5.1 Data Retrieval Information 21
5.2 Missing Data 21
5.3 Reporting conventions 22
5.4 Multiple recordings of an event for the same patient 22
5.5 Presentation of results 22

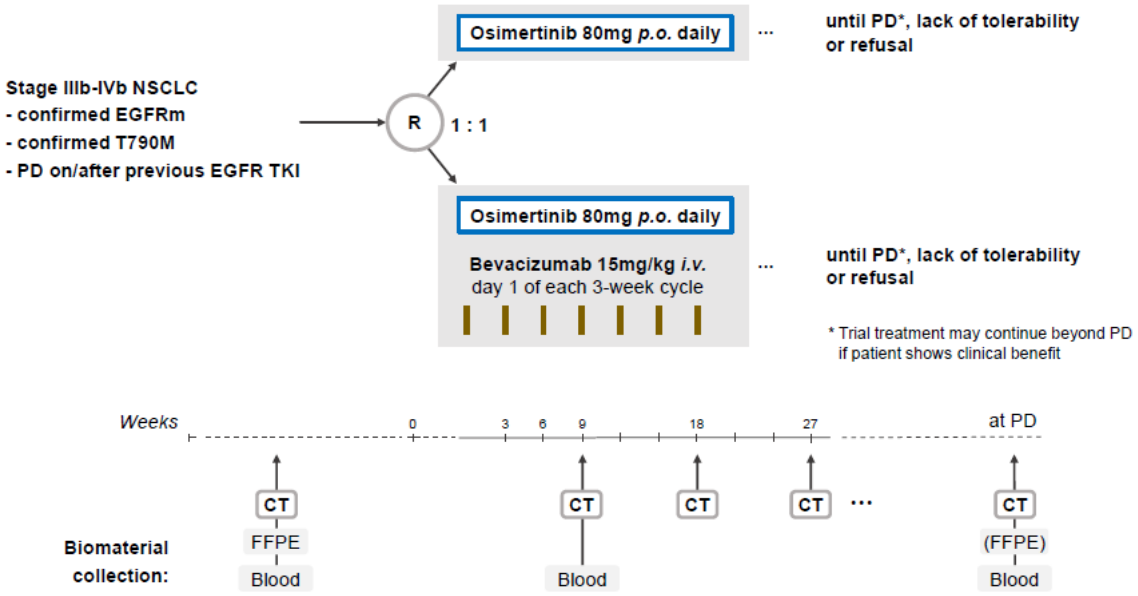
INTRODUCTION

The aim of this Statistical Analysis Plan (SAP) is to describe an analytic and solid framework that will be followed in order the final efficacy analysis of the BOOSTER trial to be implemented. A short description of the contents of this statistical analysis plan is provided below:

1. **Trial oversight**: trial's schema, objectives and trial endpoints, eligibility criteria, study treatment, statistical design (sample size and power), trial duration
2. **Statistical considerations for final analysis**: analysis timing, definition of primary and secondary endpoints, (serious) adverse events definition, analysis populations
3. **Primary efficacy analysis of progression-free survival (PFS)**
4. **Additional secondary analysis:** accrual and baseline characteristics, follow-up and treatment administration, secondary and exploratory analysis
5. **Technical issues**: data retrieval, testing, handling of missing data, reporting conventions.

1 Trial oversight

This is a randomized controlled phase II multinational, multi-center trial of osimertinib and bevacizumab versus osimertinib alone as second-line treatment in patients with stage IIIb-IVb NSCLC harbouring activating EGFR (exon 19 deletion or L858R) and T790M resistance mutation.



SCHEMA 1. Trial design

BOOSTER is a trial with block stratified randomization balanced by institution and patients randomized 1:1 to the experimental and control arm. The two stratifications factors are: ethnicity (Asian, Non-Asian) and material used for T790M testing (tumor vs circulating DNA).

1.1 Objectives

Primary objective

The **primary objective** of the study is to assess the efficacy of the combination of osimertinib and bevacizumab versus osimertinib alone in terms of progression-free survival (PFS) assessed by RECIST 1.1 criteria.

Secondary objectives

The **secondary objectives** of the study include:

- To compare short and long-term clinical efficacy outcomes
- To assess the tolerability of the two treatment

1.2 Endpoints

Primary endpoint:

- Progression-free survival (PFS) based on RECIST 1.1 criteria

Secondary endpoints:

- Objective response (OR) based on RECIST 1.1 criteria
- Disease control defined as complete or partial response, or disease stabilization confirmed at subsequent radiological assessment
- Adverse events graded according to CTCAE v4.0
- Overall survival (OS)

Correlative endpoints:

- T790M evolution in tissue and plasma/serum (between baseline and relapse).
- Monitoring of mutation burden and specific EGFR mutations (del19, L858R, T790M) in sequential plasma/serum samples
- Analysis of bevacizumab/osimertinib resistance mechanism in relapse biopsy samples (NGS)

1.3 Most important eligibility criteria

Inclusion criteria:

- NSCLC, stage IIIb/IIIc (not amenable to radical therapy) or IVa/IVb according to 8th TNM classification, after progression following prior EGFR TKI (erlotinib, gefitinib, dacomitinib or afatinib) therapy as the most recent treatment regimen.
- Pathological diagnosis of predominantly non-squamous NSCLC.
- Maximum of one line of previous platinum-based chemotherapy.
- Histological or cytological confirmation of EGFRm (exon19 deletion or exon 21 L858R).
- Locally confirmed T790M mutation determined from biopsy (preferred) or on circulating tumor DNA, documented in tissue, plasma or serum after disease progression on the most recent EGFR TKI regimen.
- Plasma, serum, and tumor (preferred) tissue or cytology (if biopsy was taken and FFPE tumor material is not yet fully depleted) after disease progression on the most recent EGFR TKI treatment available for central confirmation of T790M.
- Measurable or evaluable disease

- Adequate hematological, renal and liver function
- Performance status 0-2

Exclusion criteria:

- Patients with mixed NSCLC with predominantly squamous cell cancer, or with any small cell lung cancer (SCLC) component.
- Symptomatic or active central nervous system metastases, as indicated by clinical symptoms, cerebral edema, and/or progressive growth.
- Previous treatment with osimertinib and/or bevacizumab
- Patients currently receiving medications or herbal supplements known to be potent CYP3A4 inducers
- Any unresolved toxicities from prior therapy greater than CTCAE v4.0 grade 1

1.4 Trial treatment

Experimental arm: Osimertinib, 80 mg p.o., once daily plus bevacizumab 15 mg/kg i.v. on day 1 of every 3-week cycle.

Control arm: Osimertinib, 80 mg p.o., once daily

Patients will receive the treatment until progression, lack of tolerability, or patient declines further protocol treatment. Trial treatment may also continue beyond progression for as long as the patient may still derive benefit as per investigator decision.

1.5 Statistical design, sample size & power

The current randomized phase II trial (stratification factors: Ethnicity (Asian, Non-Asian) and material used for T790M testing (tumor vs circulating DNA)) is a superiority trial aiming to compare PFS between the two randomized arms (combination of osimertinib and bevacizumab versus osimertinib alone).

The median PFS for the target population under osimertinib alone is assumed to be 11.0 months. In order to **detect a 36% improvement in PFS** under the combination of osimertinib and bevacizumab with **80% power at the 5% one-sided significance level**, a total number of **126 PFS events is required**. For the **targeted HR=0.64, corresponding to an increase for the median PFS from 11.0 to 17.2 months**, a total of **154 randomized patients** need to be followed for an expected duration of 48 months, assuming an accrual period of 29 months (accrual is assumed to be non-linear with

increasing rate to 6 patients per month, after the first 6 months). A cumulative loss to follow-up rate of 5% by 30 months is assumed.

Thus, in this phase II trial, 77 patients will be randomized 1:1 into each arm. PFS is measured, according to RECIST 1.1 criteria, from randomization and compared between treatment arms in the ITT cohort.

An interim efficacy analysis for the primary endpoint of PFS is also included in the statistical design, which will be carried out when 63 (50%) of the 126 PFS events have been observed (expected to occur approximately 26 months after the randomisation of the first patient). According to O' Brien-Fleming approach, in this interim analysis, the treatment effect on PFS will be tested at one-sided type I error rate of 0.6%. If the boundary is crossed in favour of the alternative hypothesis of superiority, the recruitment will stop early and a significant PFS benefit will be claimed for the combination. Otherwise, the trial will proceed to full accrual and the final analysis (with one-sided type I error rate of 4.4%) will be performed when the total number of 126 PFS events has been observed.

1.6 Total trial duration

The total trial follow-up duration to observe the required events is expected to be 48 months with an interim efficacy analysis based on O' Brien-Fleming boundary at 26 months. Taking into account a run-in period of 6 months and an additional 6 months for the final analysis report, the total trial duration is expected to be 5 years from randomization of first patient.

End of trial occurs when both of the following criteria have been satisfied:

a) The trial is mature for the analysis of the primary endpoints as defined in the protocol
b) The database has been fully cleaned and frozen for this analysis.

2 Statistical considerations for final analysis

2.1 Analysis timing

According to the statistical design, the final analysis will be carried out when the 126 required PFS events will be available for the BOOSTER randomized patients. This is expected to occur approximately 48 months after the randomization of the first patient, assuming an accrual period of 29 months.

2.2 Study's endpoints

2.2.1 Primary endpoint

The primary endpoint of the trial is PFS, defined as the time from the date of randomization until documented progression (based on RECIST 1.1 criteria) or death, if progression is not documented. Censoring for PFS (patients without progression/death) will occur at the last tumor assessment. In the frame of a sensitivity analysis, if the last tumor assessment is "Non evaluable" (NE), censoring will occur to the most recent tumor assessment where an overall evaluable result is recorded.

2.2.2 Secondary endpoints

Secondary endpoints, according to the protocol, include OR, disease control, OS and adverse events (AEs). More specifically:

- OR is defined as the best overall response (complete or partial) across all assessment time-points according to RECIST 1.1 criteria, during the period from randomization to termination of trial treatment. In the frame of a sensitivity analysis, all timepoints, until the end of follow-up for progression, will be taken into account for the determination of best overall response.
- Disease control is defined as complete or partial response, or disease stabilization, confirmed at subsequent radiological assessment
- OS is defined as the time from the date of randomization until death from any cause. Censoring for OS (patients without death) will occur at the last follow-up date
- Toxicity, defined as AEs graded according to CTCAE v4.0.

2.2.3 Exploratory endpoints

- Time to treatment failure (TTF) is defined as time from the date of randomization until discontinuation of protocol treatment for any reason (including progression of disease, death, discontinuation of at least one of the drugs consisting the treatment combination due to any reason, such as treatment toxicity or investigator's decision, withdrawal/lost to follow-up (LFU)). Censoring for TTF (patients on treatment) will occur at the last follow-up date.

- Similar to TTF, time to treatment discontinuation (TTD) for each drug separately (osimertinib or bevacizumab) is defined as the time from the date of randomization until discontinuation of the specific protocol drug for any reason.
- Duration of response (DoR) is defined as the time from the documentation of tumor response (either partial or complete response) to disease progression or death. Censoring will occur at the last tumor assessment with response other than progression.
- Duration of clinical benefit (DoCB) is defined as the time from documentation of clinical benefit (stable disease, partial or complete response) to disease progression or death. Censoring will occur at the last tumor assessment with response other than progression.

2.2.4 Correlative endpoints

Correlative endpoints of the study include T790M evolution in tissue and plasma/serum (between baseline and relapse), monitoring of mutation burden and specific EGFR mutations (del19, L858R, T790M) in sequential plasma/serum samples and analysis of bevacizumab/osimertinib resistance mechanism in relapse biopsy samples (NGS). The statistical analysis of these correlative endpoints will be specified in a separate SAP document.

2.3 (Serious) Adverse Events

Adverse events (AE)

The main criterion for treatment tolerability is the occurrence of toxicities and adverse events. The severity and causality will be classified according to the NCI CTCAE v4.0. The CTCAE v4.0 is available for downloading (http://evs.nci.nih.gov/ftp1/CTCAE/About.html).

An AE is defined as any untoward medical occurrence that occurs from the date of signature of inform consent until 30 days after all trial treatment discontinuation, regardless of whether it is considered related to a medication. In addition, any known untoward event that occurs subsequent to the adverse event reporting period that the investigator assesses as possibly related to the protocol treatment should be considered an adverse event.

Serious Adverse Events (SAE)

An SAE is defined in general as any undesirable medical occurrence/adverse drug experience that occurs during or within 30 days after stopping study treatment that, at any dose:

- results in death (any cause, except progression of cancer under study)
- is life-threatening
- requires or prolongs inpatient hospitalization
- results in persistent or significant disability/incapacity
- is a congenital anomaly or birth defect (including neonatal deaths)
- is a secondary malignancy
- is an event of clinical interest (drug induced liver injury, overdose)

Other significant/important medical events which may jeopardize the patient are also considered serious adverse events. Serious also includes any other event that the investigator or the ETOP Safety Office judges to be serious or which is defined as serious by the regulatory agency in the country in which the event occurred.

Severity Grade of (serious) adverse event

The adverse event severity grade provides a qualitative assessment of the extent or intensity of an adverse event, as determined by the investigator or as reported by the subject. *The severity grade does not reflect the clinical seriousness of the event*, only the degree or extent of the affliction or occurrence (e.g., severe nausea, mild seizure), and does not reflect the relationship to study drug. A severe event may be of relatively minor medical significance (such as severe headache).

Severity grade for other adverse events not covered in the toxicity grading scale:

Grade 1	Mild- transient or mild discomfort; no limitation in activity; no medical intervention/therapy required
Grade 2	Moderate- mild to tolerate limitation in activity; some assistance may be needed; no or minimal medical intervention/therapy required
Grade 3	Severe- marked limitation in activity, some assistance usually required; medical intervention/therapy required, hospitalization is possible
Grade 4	Life-threatening- extreme limitation in activity, significant assistance required; significant medical intervention/therapy required, hospitalization or hospice care probable
Grade 5	Fatal- the event results in death

Causality of (serious) adverse event

The investigator must determine the relationship between the administration of trial drug(s) and the occurrence of an AE/SAE following the definitions indicated below:

- Not suspected (unrelated/unlikely): The temporal relationship of the adverse event to trial drug(s) administration makes a causal relationship unlikely or remote, or other medications, therapeutic interventions, or underlying conditions provide a sufficient explanation for the observed event.
- Suspected (possible/probable/definite): The temporal relationship of the adverse event to trial drug(s) administration makes a causal relationship possible, and other medications, therapeutic interventions, or underlying conditions do not provide a sufficient explanation for the observed event.

2.4 Analysis populations

Efficacy cohort: The efficacy analysis population includes all randomized subjects. Patients who were randomized but did not receive trial treatment will be also included in the efficacy cohort (intention-to-treat (ITT) population).

Safety cohort: The safety analysis population consists of all subjects who have received at least one dose of trial treatment according to the treatment they actually received, regardless of their allocated treatment at randomization.

Of note, there will be also an early safety evaluation after the first 10 patients have been randomized to the experimental arm (total 20 patients, 10 in the experimental and 10 in the control arm) and

have received trial treatment for 30 days. This safety evaluation will be submitted to the ETOP Independent Data Monitoring Committee (IDMC) for advice. The recruitment into the trial will continue while safety is evaluated.

3 Primary efficacy analysis of progression-free survival (PFS)

Primary efficacy analysis of PFS (primary endpoint) will be performed on all randomized patients, based on their initial treatment assignment (ITT, efficacy cohort). PFS time is measured from randomization.

Formal hypothesis testing

The study is designed to test the hypothesis that treatment combination of osimertinib and bevacizumab will lead to an increase in median PFS to 17.2 months, from 11.0 months under standard treatment with osimertinib alone. **According to the study design, this corresponds to a HR of 0.64**. Using **80% power** and a **one-sided type I error of 5%,** a total of **126 PFS events** are needed to be observed in order to achieve the trial goal, while an **interim analysis** is scheduled to be performed at 50% of the information time, i.e., when **63 PFS events** have been observed. To keep the overall one-sided type I error at 5% according to the O' Brien-Fleming approach (calculated by the Lan-DeMets spending function), a one-sided alpha 0.6% is assigned to the designed interim efficacy analysis, while the exact p-value boundary for the final efficacy analysis is 4.8%.

In the frame of final efficacy analysis, the **formal comparison of the PFS between the two treatment arms**, will be based on the **stratified log-rank test** (with ethnicity and material used for T790M testing being the stratification factors). Unstratified log-rank test will be also calculated. This would be of particular value in the case at least one stratum level has very low number of patients.

Further PFS analyses

The following PFS analyses will be also performed and presented

- The total number (%) of PFS events observed, overall and by treatment arm will be presented. In addition, 1/2-year PFS estimates, median PFS and respective 95% CIs will be provided. Respective results by stratification factors will be also presented.
- Graphical representation of PFS, by treatment arm will be performed via a Kaplan-Meir plot. The plot will be also produced separately by stratum.
- Number of PFS events, median PFS and unstratified/unadjusted HRs (along with 95% CIs), interaction p-value between treatment and each variable of interest, will be summarised for the subgroups defined by treatment and the following variables of interest: stratification factors (ethnicity, material for T790M testing), gender, age (appropriately categorized), smoking history, ECOG performance status at diagnosis, stage of tumor, EGFR mutation subtype and biological markers (if available). This information will be depicted in a tabular format in the report and a forest plot will be produced for the publication/presentation.
- A table with information about the sites of progression will be also provided.

- Furthermore, to assess the effect of trial treatment and other clinicopathological variables on PFS, Cox proportional hazards model will be fitted.
    - Initially a univariate (stratified and unstratified) Cox model will be fitted in the model and the statistical significance of trial treatment will be tested at the 5% significance level.
    - Subsequently, multivariate Cox models (stratified and unstratified) will be estimated, adjusted for the clinicopathological variables of interest as defined above (in case of unstratified model, the stratification factors will be also included as covariates in the model). The backward elimination method, with a removal criterion at 10% will be implemented to conclude on the statistically significant variables of the model. The HRs along with the corresponding 95% CIs for all significant predictors (in the multivariate Cox model) will be summarised in a tabular format in the report and the corresponding forest plot will be subsequently produced for presentation/publication.
    - The proportionality assumption of Cox models will be explored by Schoenfeld's residuals and by testing for time-dependent effect of covariates in extended Cox models. In cases that non-proportionality is detected further appropriate measures will be used:
        - Use of variable(s), for which proportionality assumption is violated, as stratification factor(s)
        - Use of weighted tests, alternatives to log-rank for the comparison of survivals, such as the Wilcoxon test
        - Estimation of Restricted Mean Survival Time (RMST) at specific time points (close to median follow-up and covering the full follow-up time for the majority of patients)
        - Calculation of HR for separate time intervals

4 Additional secondary analysis

In this section, detailed information about the additional analysis that will be performed in the frame of final efficacy analysis for BOOSTER trial is presented.

4.1 Patient accrual, balance of stratification factors and baseline characteristics

- Patient accrual by center and country will be presented in tabular format.
- In addition, expected vs. observed accrual will be graphically displayed.
- For patients deemed ineligible (patients registered in the online database but eventually not randomized) a table summarizing the reasons for non-randomization will be provided.
- Balance of treatment allocation by center and by stratification factor will be summarised as well.
- Patient & tumor baseline characteristics (categorical: ethnicity, gender, smoking history, ECOG performance status at diagnosis, tumor stage, material used for T790M testing, EGFR mutation subtype and continuous: age at randomization), will be presented overall and separately by treatment arm. Frequencies and corresponding percentages will be presented for categorical variables (if missing cases exist, a separate category named "*Missing*" will be created), while the following descriptive measures will be considered for the continuous ones: n (non-missing sample size), mean, 95% CI for the mean, median, maximum and minimum. Balance of baseline characteristics by treatment arm will be assessed via the Fisher's exact test for categorical variables and the Mann-Whitney U test for continuous.
- Also available information on medical history and prior treatment will be summarised, by treatment arm.

4.2 Follow-up information and treatment administration

Firstly, a consort flow diagram will be created to graphically depict the progress through the phases of the trial.

Median follow-up (FU) of the patients (overall and by treatment arm) along with the respective interquartile range (IQR) and the number (%) of patients that are still alive, will be summarised in a table. A Kaplan-Meier plot, overall and by treatment arm, will be also provided for a graphical representation of the respective information.

Treatment information will be summarised overall and separately by treatment arm. More specifically the following information will be presented:

- Number of patients that started treatment, information on number of cycles (median, min-max). For those patients randomized, progressed but with physician's and their own agreement continued receiving treatment, information on treatment cycles as well as treatment failure after 1st progression will be additionally provided.
- Number of patients that did not receive any dose of trial treatment, along with reasons for not doing so
- Number of treatment failures, 1/2-year TTF estimates and median TTF time along with the corresponding 95% CIs and the reasons for treatment failure/discontinuation will be presented overall and by treatment arm. The stratified log-rank test will be used to compare TTF between the two treatment arms.
- A Kaplan Meier plot for TTF, by treatment group, will be created.
- For patients with treatment failure, information for further lines of treatment will be also provided.

4.3 Secondary (efficacy) and exploratory endpoints

4.3.1 Overall survival

Similar to PFS the following will be presented for OS:

- Total number (%) of OS events, overall and by treatment arm (as well as by stratification factor)
- 1/2-year OS estimates, median OS and respective 95% CIs (comparison between the arms based on stratified and unstratified log-rank test)
- Kaplan- Meier plot by treatment arm and separately for each stratum
- Subgroup analysis (number of deaths, median OS and unstratified/unadjusted HRs (along with 95% CIs)) by treatment and variables of interest
- Univariate and multivariate Cox proportional hazards model, adjusted for the stratification factors and clinicopathological variables of interest; HRs and corresponding 95% CIs for all significant OS predictors
- Summary table of the death causes (e.g., lung cancer, toxicity, other, etc) overall and by treatment arm.

Of note, at the time of final analysis for the primary endpoint (PFS), the OS events are not expected to be sufficient to provide enough power for comparison between treatments. In case that the endpoint is immature at that timepoint an additional OS analysis with longer follow-up will be performed at a later time point.

4.3.2 Objective response rate & disease control rate

- Best overall responses (BOR) as well as objective response rate (ORR) and disease control rate (DCR) will be presented overall and separately for the two treatment arms, along with a 95% exact binomial CI.
- ORR and DCR will be compared between the two treatment groups using Fisher's exact test and Cochran-Mantel-Haenszel test stratified by the stratification factors of the trial.
- Logistic regression models will be further applied to investigate the treatment effect, adjusting for stratification factors and variables of clinical interest.
- A waterfall plot will be created by treatment arm to present the best percent change in tumor size (sum of target lesions diameter) from the baseline tumor assessment before randomization
- The percent changes in tumor size (sum of target lesions diameter) from the baseline tumor assessment before randomization over the time will be depicted graphically by a spider plot, separately for each treatment arm.

4.3.3 Duration of response and duration of clinical benefit

- Median DoR and DoCB, along with the corresponding 95% CIs will be presented, for all responders and separately for the two treatment groups.
- Duration of response and clinical benefit (including patients with at least stable disease) will be compared between the two treatment arms using Kaplan-Meier method.
- Graphical representation of DoR and DoCB will be also performed via swimmer plots, separately for each treatment arm.

4.3.4 Subgroup analysis

To determine whether the treatment effect is consistent across various subgroups the between-group treatment effect for all efficacy endpoints will be estimated within each category of the following pre-specified variables.

<u>Main subgroup analyses:</u>

- Ethnicity
- Stage
- Gender

<u>Other pre-planned subgroup analyses:</u>

- Material used for T790M testing
- EGFR mutation subtype
- Tumor Mutation Burden at baseline (if available)

- Age group

Notes:

1. In case of imbalance in number of patients with the subgroups created by treatment and the variables of interest, this subgroup analysis will not count in the multiple comparison adjustment.
2. The interaction of each subgroup with the treatment effect on PFS and OS will be tested based on respective Cox models.

4.4 Sensitivity efficacy analysis

In a sensitivity analysis framework, the efficacy analysis will be repeated using the safety cohort.

4.5 Safety analysis

The safety analysis will be performed based on the safety cohort (i.e., patients who have received at least one dose of trial treatment) and will include the following:

- Overview of the number of patients who experienced an AE/SAE, as well as the number of patients in the safety cohort who did not experience an event, along with respective percentages will be shown. This information will be presented overall and by treatment arm. Also, number of patients that entered the study with baseline symptoms will be reported.
- Number of AEs/SAEs and rate of occurrence per month of follow-up, overall and by treatment arm.
- Number of patients experiencing a specific number of AEs/SAEs, overall and by treatment arm.
- Distribution of AEs/SAEs by grade and CTCAE category, overall and separately for the two treatment arms. Six columns, one for each grade and one for all (any) grades, will be shown (for each arm). An additional column (by arm) indicating which events were SAEs -or started as AEs and became SAEs later on- will be also available. In this column, the frequency of the SAEs and the severity grade will be given. The percentages that will accompany the frequencies will be based on the respective frequency of an event over the total number of patients in the safety cohort and specific treatment arm. This table will include all AEs/SAEs irrespective of their relation to the trial treatment.
- Analogous table focusing only on the treatment related AEs/SAEs.
- Number and corresponding percentages of treatment related AEs/SAEs, leading either to treatment discontinuation or death will be summarised for the two treatment arms and overall.
- Treatment related AEs/SAEs (of any grade) occurring in more than 10% (or any other relevant %) will be presented for the two treatments.

- The risk difference, along with corresponding 95% CIs for specific AEs/SAEs (for example most frequent/related events (i.e., ≥10%) or events of grade ≥3), between the two treatment arms will be presented and graphically depicted.
- Maximum severity of AEs/SAEs per patient, overall and by treatment arm.
- Number of SAEs by center.
- For all fatal SAEs, cause of death will be provided.

5 Technical details

Data will be primarily analysed using the SAS software package (version 9.4), while the R statistical software will be also used for specific analyses and plots.

All final analysis and reviews will be performed according to the Standard Operating Procedures (SOPs) of the Frontier-Science Foundation Hellas (FSFH) statistical team. A second statistician, the reviewing statistician, will independently reproduce all analysis and summary statistics. The reviewing statistician will have an overview of the entire analysis and will explicitly check the code producing tables and figures, as well as any other pieces of code as desired.

5.1 Data Retrieval Information

The final analysis will be based on the database download that will take place, as soon as the total number of 126 PFS events required according to the statistical design of the trial are observed. Using this database extraction, a set of queries will be produced and forwarded to trial's data manager with the expectation to be answered in a pre-specified time period (approximately four weeks). Corrections and responses based on these queries, will be used for correcting the previously downloaded database, in order to create the final clean dataset to be used for the analysis.

5.2 Missing Data

Baseline characteristics

For categorical baseline characteristics if missing cases exist, a separate category named 'Missing' will be created. As far as continuous values, missing cases will not be replaced by any statistics calculated over non-missing data.

Dates

If the day of the month is missing for any date used in the analysis, the 15th of the month will be used to replace the missing date unless the calculation results in a negative time duration (e.g., date of onset cannot be prior to day one date). In this case, the date resulting in one day of duration will be used. If the day of the month and the month are missing for any date used in a calculation, January 1 will be used to replace the missing date. Missing dates for adverse events will be imputed based on the similar principle.

Incomplete tumor assessment information

In patients who have no on-study assessments:

- If death is recorded prior to the first planned tumor assessment, the death date will be considered as the date of the PFS event.

- If clinical progression is recorded prior to the first planned tumor assessment, the date of the reported clinical progression will be considered as the date of the PFS event.
- In all other cases, the patient will be censored at the date of randomization plus 1 day.

5.3 Reporting conventions

Regarding the estimates presented in the report, the following rules will be adopted:

- P-values ≥0.001 will be reported with three decimal places
- P-values >0.010 will be reported with two significant decimal digits
- P-values less than 0.001 will be reported as '<0.001'
- Means, medians, 95% confidence intervals (CIs), quantiles, and any other statistics, will be reported with one decimal digit
- Hazard ratios (HRs) and their 95% CIs will be reported with two decimals
- Estimated parameters, not on the same scale as raw observations (e.g., regression coefficients) will be reported with three significant figures

5.4 Multiple recordings of an event for the same patient

There are some cases where a patient may experience the same event (AE/SAE) more than one time. In such cases, the event will be count only once (with the highest grade) for the calculation of the total number of events.

5.5 Presentation of results

The results will be presented through tables and figures. A summary of the results will also accompany the main report. First, a short synopsis of the results will be presented through bullets, where only the most important findings will be shown. Following that, a more detailed description of the results will be provided, sectioned in the following order:

I. Patient accrual and baseline characteristics
II. Follow-up and treatment administration
III. Efficacy analysis
IV. Safety analysis

All tables and figures will be included in an appendix.