ETOP 15-19 ABC-lung

A randomized non-comparative open label phase II trial of atezolizumab plus bevacizumab, with carboplatin-paclitaxel or pemetrexed, in EGFR-mutant non-small cell lung carcinoma with acquired resistance

ABC-lung: Atezolizumab, Bevacizumab and Chemotherapy in EGFR-mutant non-small cell lung carcinoma

Statistical Analysis Plan (SAP) for Final Analysis

A clinical trial of ETOP IBCSG Partners Foundation

Protocol version 2.0 (20210112)

SAP VERSION	v1.0
SAP DATE	12 October 2023
TRIAL STATISTICIANS	Urania Dafni, Katerina Vervita
TRIAL PRIMARY INVESTIGATORS	Ross Soo, Margarita Majem Tarruella, Delvys Rodriguez Abreu, Martin Früh
FINAL ANALYSIS SAP AUTHORS	Urania Dafni, Katerina Vervita, Zoi Tsourti

Contents

INTRODUCTION

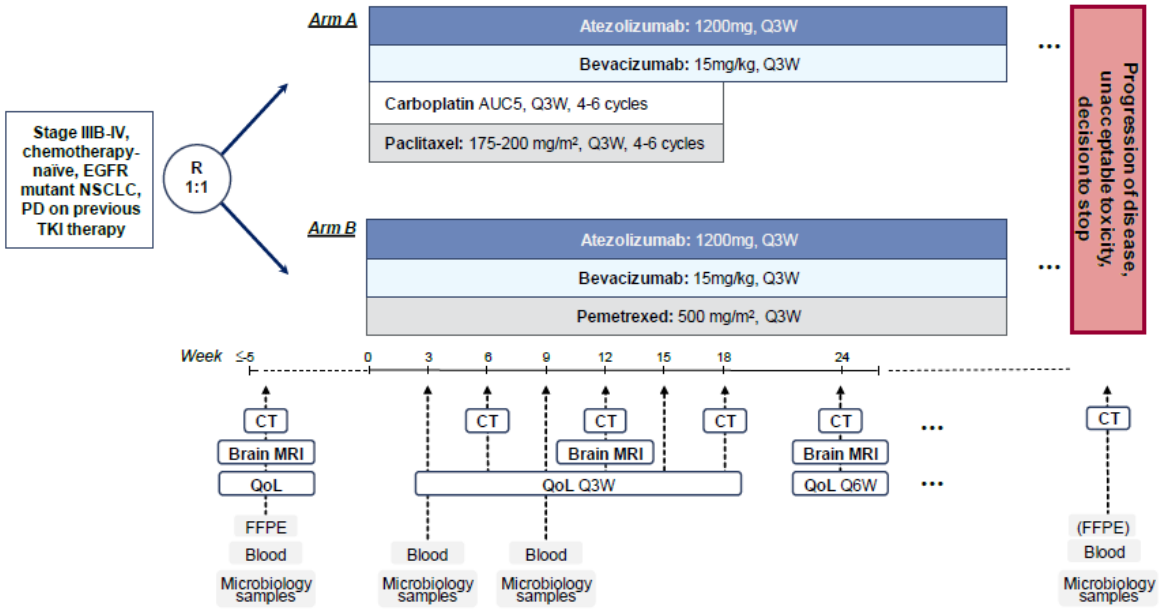
The aim of this Statistical Analysis Plan (SAP) is to describe an analytic and solid framework that will be followed in order for the **final efficacy analysis** of the ETOP 15-19 ABC-lung trial to be implemented (based on protocol version 2.0 (20210112)).

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 secondary and exploratory 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.
6. **List of tables and figures**

1 Trial oversight

This is a multinational, multi-centre, randomized, non-comparative open label phase II trial of atezolizumab plus bevacizumab, with carboplatin-paclitaxel or pemetrexed, in epidermal growth factor receptor (EGFR)-mutant non-small cell lung carcinoma (NSCLC) with acquired resistance.



SCHEMA 1. Trial design

Target Population: Patients with chemotherapy-naïve, immune checkpoint inhibitor-naïve, EGFR-mutant (L858R or del19) stage IIIB/C (not amenable to radical therapy) or IV non-squamous NSCLC that have relapsed after 1-2 lines of EGFR tyrosine kinase inhibitor (TKI) for metastatic disease. Patients with T790M mutation must be previously treated with a third-generation TKI (e.g., osimertinib).

ABC-lung is a trial with block stratified randomization (blocks of random size, multiples of two) balanced by institution. Patients randomized 1:1 to one of the two treatment arms: atezolizumab, bevacizumab and carboplatin plus paclitaxel (Arm A) or atezolizumab, bevacizumab and pemetrexed (Arm B). The stratifications factor is: prior third-generation TKI (e.g., osimertinib) versus no prior third-generation TKI.

Although randomization is used to allocate patients to either arm A or arm B, the study is not powered nor designed to be comparative in nature. The purpose of randomization is to reduce bias due to patient selection into either treatment arm.

1.1 Previous protocol versions

The first protocol version of ABC-lung trial was released on October 31st, 2019, while a protocol amendment (v2.0) has been approved on January 12th, 2020. The main changes that occurred in the amended protocol have been the following:

- Due to increased haematological toxicities observed in Asian patients in the IMpower150 trial, it is recommended that the starting dose of paclitaxel in arm A should be 175 mg/m^2 every three weeks instead of 200 mg/m^2 in the original protocol.
- Patients should receive anti-emetics according to the local standard of care and manufacturer's instruction. Corticosteroids as a premedication for chemotherapy is permitted and is dosed according to local practice.
- Regarding exclusion criteria:
    - Patients are not allowed to receive any biologic drugs targeting the immune system (for example, TNF blockers, anakinra, rituximab, abatacept, or tocilizumab) as in the first protocol version, unless 6 weeks have passed prior to treatment start.
    - Patients should not have history of active diverticulitis.
    - Core biopsy or other minor surgical procedure, excluding placement of a vascular access device, within 7 days prior to the first dose of bevacizumab is forbidden as in the original protocol. In addition, based on the amendment protocol, major surgery or significant traumatic injury within 28 days prior to the first dose of bevacizumab or minor surgical procedure within 7 days, or placement of a vascular access device 2 days prior to the first dose of bevacizumab are not allowed.

All patients have been recruited under the protocol amendment (v2.0).

1.2 Objectives

Primary objective

The **primary objective** of this study is to explore the clinical efficacy of atezolizumab and bevacizumab combined with chemotherapy in terms of progression-free survival (PFS), in patients with EGFR-mutant advanced NSCLC after failure of standard EGFR TKIs.

Secondary objectives

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

- To evaluate secondary measures of clinical efficacy including objective response rate (ORR), extra-cranial PFS, intra-cranial PFS and overall survival (OS)
- To assess the safety and tolerability of the treatment.

1.3 Endpoints

Primary endpoint:

- PFS rate at 12 months according to RECIST v1.1

Secondary endpoints:

- Objective response (OR) according to RECIST v1.1
- Extra-cranial PFS
- Intra-cranial PFS
- Overall survival, including OS rate at 12 months
- Adverse events (AEs) graded according to CTCAE v5.0
- Patient reported Quality of Life (QoL)

Exploratory endpoints:

- Subgroup analysis according to
    - EGFR subtype
    - prior use of third-generation TKI or not
    - PD-L1 expression levels
- DNA from tumour and blood samples will be isolated and used for sequencing of specific gene panels and, if feasible, tumour mutation burden. Microbiome analysis will be performed on oropharyngeal swabs and faecal samples

1.4 Most important eligibility criteria

Inclusion criteria at randomization:

- Chemotherapy naïve, non-squamous NSCLC, stage IIIB/C (not amenable to radical therapy) or IV. Patients who have received previous adjuvant or neoadjuvant chemotherapy are eligible if the date of last dose of treatment was at least 12 months before randomization

- Known EGFR mutations genotypes by tissue or ctDNA; patients with common mutations (L858R or Del19) and other rare mutations (e.g., S768I, G719X) are eligible
- Measurable or evaluable disease by RECIST v1.1
- Disease progression (during or after) or unacceptable side effects from prior treatment with at least one EGFR TKI (TKI washout period = 7 days).

  If most recent line of treatment (1st or 2nd line) was a third-generation EGFR TKI (e.g. osimertinib):

  - Patient must be known to be EGFR mutation positive, either on fresh tumour biopsy taken >7 days prior to protocol treatment start or by recent ctDNA analysis (informative ctDNA test, local test).
  - T790M genotype is allowed.

  If most recent line of treatment (1st or 2nd line) was a first- or second-generation EGFR TKI (e.g. afatinib, dacomitinib, erlotinib, gefitinib):

  - Patient must be known to be tissue EGFR T790M wild type (local test) on most recent line of EGFR TKI or if no tissue re-biopsy, no evidence of T790M on ctDNA but identified L858R, del19, S768I or G719X genotypes (informative ctDNA test, local test).
- Treatment with an EGFR TKI therapy for at least 30 days
- Adequate haematological, renal (CrCl at least 45ml/min) and liver function
- Willing to make available surplus tissue obtained at the time of acquired resistance to EGFR TKI

<u>Exclusion criteria</u> at randomization:

- Prior systemic cytotoxic chemotherapy for advanced stage NSCLC
- Prior therapy with bevacizumab or other anti-angiogenic agent
- Prior immune checkpoint inhibitor therapy
- More than two lines of EGFR TKI therapy
- Known small-cell lung carcinoma (SCLC) or high grade neuroendocrine carcinoma (if progression biopsy has been performed locally)
- Squamous cell histologic subtype
- Known EGFR T790M positive genotype by tissue on most recent EGFR TKI progression or ctDNA and have not received an approved EGFR TKI targeting T790M

- Active or untreated CNS metastases as determined by brain MRI
- Patients with CNS metastases must be non-progressive by RECIST v1.1 and symptomatically stable with no ongoing requirement for corticosteroids as therapy for CNS disease; anticonvulsants at a stable dose allowed.
- Radiotherapy in target lesions within 4 weeks of randomization
- QTc of grade ≥3 according to CTCAE v5.0
- Active autoimmune disease that has required systemic treatment in past 2 years
- Active or uncontrolled HIV, tuberculosis, hepatitis B or C infection
- Inadequately controlled hypertension (defined as systolic blood pressure >150 mmHg and/or diastolic blood pressure >100 mmHg). Anti-hypertensive therapy to achieve these parameters is allowable.
- Prior history of hypertensive crisis or hypertensive encephalopathy
- Significant vascular disease (e.g. aortic aneurysm requiring surgical repair or recent peripheral arterial thrombosis) within 6 months prior to randomization
- History of haemoptysis (≥2.5 ml of bright red blood per episode) within 1 month prior to randomization
- Recent surgery: Core biopsy or other minor surgical procedure, excluding placement of a vascular access device, within 7 days prior to the first dose of bevacizumab.
- Serious, non-healing wound, active ulcer, or untreated bone fracture
- Proteinuria, as demonstrated by urine dipstick or >1.0 g of protein in a 24-hour urine collection
- Any unresolved toxicities from prior therapy greater than CTCAE v5.0 grade 1 at the time of starting trial treatment with the exception of alopecia

1.5 Trial treatment

Arm A:

- Atezolizumab (1200 mg) Q3W, until PD*
- Bevacizumab (15 mg/kg), Q3W, until PD
- Carboplatin (AUC5) Q3W, 4-6 cycles
- Paclitaxel‡ (175-200 mg/m^2, at the investigators' discretion), Q3W, 4-6 cycles

Arm B:

- Atezolizumab (1200 mg), Q3W, until PD*
- Bevacizumab (15 mg/kg), Q3W, until PD

- Pemetrexed (500 mg/m^2), Q3W, until PD

Treatment will continue until disease progression, toxicity, or patient/physician decision.

(*) Atezolizumab treatment beyond RECIST v1.1-defined progression will be allowed if patient is continuing to derive clinical benefit.

(‡) Asian population: Due to increased haematological toxicities observed in Asian patients in the IMpower150 trial, it is recommended that the starting dose of paclitaxel should be 175 mg/m^2 every three weeks.

1.6 Statistical design, sample size & power

ABC-lung is a selection design, randomized trial with two non-comparative parallel arms.

Independently in each arm, the following hypothesis will be tested:

- H_0: 12-month PFS rate (π_0) ≤0.18, versus
- H_1: 12-month PFS rate (π_1) >0.18, evaluated at π_1=0.37.

A sample size of 45 evaluable patients (per arm) will provide a **power of 83%**, for testing the above efficacy hypothesis, using an **exact test for a single proportion**, at the **one-sided alpha of 0.025** (attained 0.023). Assuming 5% non-evaluable patients, **the required total number of randomized patients for both arms increases to 95**.

Sample size and power calculations for each arm separately have been performed using the EAST version 6.4 software.

1.7 Total trial duration

Clinical visits (until the timepoint of primary analysis) are expected to span approximately 24 months after randomization of the first patient, assuming an accrual rate of 1-2 patients per month during the first 6 months as the trial is activated by the participating centres and 12-15 patients per month thereafter (total accrual time of 12 months) plus 12 months of follow-up for all randomized patients. The duration of individual patients' trial participation is anticipated to be between 1 and 2 years.

The primary analysis will be available approximately 2.5 years after the inclusion of the first patient, that is 6 months after the last randomized patient has completed 12 months of follow-up.
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 45 evaluable patients (per arm) reach one year of follow-up time from randomization into the trial or have experienced a PFS event (progression and/or death) earlier.

2.2 Study's endpoints

2.2.1 Primary endpoint

The PFS rate at 12 months is the primary endpoint of this trial. It is defined as the rate of patients without a PFS event at 12 months from randomization. PFS is defined as the time from the date of randomization until documented progression (according to RECIST v1.1) or death, if progression is not documented. Censoring (for patients without documented progression or death) will occur at the last tumour assessment. Patients without any post-baseline tumour assessment will be censored at the date of randomization (plus 1 day).

Of note, progression based on symptomatic deterioration only, without any formal tumor assessment does not comply with the formal RECIST v1.1 definition of documented disease progression. In these cases, any subsequent formal tumor assessment with reported progression or the death date (if not subsequent tumor assessment exists) will be considered as the date of PFS event. recorded. A sensitivity analysis will be performed using reported clinical deterioration as PFS event.

As additional 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, extra-cranial PFS, intra-cranial PFS, OS, toxicity and QoL. More specifically:

- OR is defined as the best overall response (complete response (CR) or partial response (PR)) according to RECIST v1.1, across all assessment timepoints, from randomization until either (i) the end of protocol treatment or (ii) the end of follow-up.
- Extra-cranial PFS is defined as the time from the date of randomization to documentation of disease progression outside the central nervous system (CNS) as

per RECIST v1.1 or death, whichever occurred first. Censoring (for patients without documented extra-cranial progression or death) will occur at the last tumour assessment. Only intra-cranial progression is considered to be competing risk. Patients without any post-baseline tumour assessment will be censored at the date of randomization (plus 1 day).

- Intra-cranial PFS is defined as the time from the date of randomization to first documented radiographic evidence of CNS progression or death. CNS progression is defined as progression due to newly developed CNS lesions and/or progression of pre-existing baseline CNS lesions. Censoring (for patients without documented intra-cranial progression or death) will occur at the last tumour assessment. Only extra-cranial progression is considered to be competing risk. Patients without any post-baseline tumour assessment will be censored at the date of randomization (plus 1 day).
- OS is defined as the time from the date of randomization until death from any cause. Censoring (for patients without documented death) will occur at the last follow-up date. Patients without any post-baseline information will be censored at the date of randomization (plus 1 day).
- The toxicity profile of the protocol treatment is evaluated in terms of AEs (any-cause as well as treatment-related) graded according to CTCAE v5.0.
- QoL is assessed by the European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30) and the lung cancer-specific module (QLQ-LC13). The key QoL outcome is the time to deterioration (TTDet) in the QLQ-C30 global health status/global QoL. TTDet is defined as the time from baseline to the first time that the patient's score shows a ≥10-point increase/decrease (higher scores for symptoms indicate worse condition, while higher scores for function scale indicate better condition) above baseline in the EORTC-QLQ-C30 global health/ QoL scale and QLQ-LC13 symptoms/subscales (cough, dyspnoea, and chest pain). Patients with no definitive deterioration events will be censored at the date of the last available QoL assessment. The primary outcome will be the global health/ QoL scale. In order for a scale/symptom to be considered "deteriorated," a score increase of ≥10 points above baseline must be held for at least two consecutive assessments or an initial score increase of ≥10 points is followed by death within 3 weeks from the last assessment.

In addition, Duration of Response (DOR), Time-to-Treatment Failure (TTF) and drug-specific Time-to-Treatment Discontinuation (TTD) will be analysed as exploratory endpoints:

- DoR is defined as the time from the date of first documentation of objective response (CR or PR, according to RECIST criteria version 1.1) to the date of first documented progression or death. Censoring will occur at the last tumour assessment with response other than progression. Patients without tumour assessment after documented objective response will be censored at the date of randomization (plus 1 day).
- TTF is defined as the time from the date of randomization to permanent discontinuation of at least one of the drugs consisting the protocol treatment due to any reason, such as toxicity, investigator decision or discontinuation due to other reasons (including progression, death or withdrawal/lost to follow-up (LFU)). Censoring for TTF (patients on treatment on all study drugs) will occur at the last follow-up date.
- Similar to TTF, TTD for each treatment drug separately (bevacizumab, atezolizumab, pemetrexed) is defined as the time from the date of randomization until discontinuation of the specific protocol drug for any reason (toxicity, investigator decision, patient refusal, progression, death, withdrawal/ LFU). Censoring for TTD will occur at the last date that patient received treatment with the specific drug.

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 CTCAE v5.0. The CTCAE v5.0 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 randomization until 90 days after the last dose of protocol treatment, regardless of whether it is considered related to a medication.

An AE can therefore be any of the following:

- Any unfavourable and unintended sign (including clinically significant abnormal laboratory finding), symptom, or disease temporally associated with the use of a protocol treatment, whether considered related to the protocol treatment or not.
- Any new disease or exacerbation of an existing disease (a worsening in the character, frequency, or severity of a known condition).
- Recurrence of an intermittent medical condition (e.g. headache) not present at baseline.

- Any deterioration in a laboratory value or other clinical test (e.g. ECG, X-ray) that is associated with symptoms or leads to a change in protocol treatment or concomitant treatment or discontinuation from protocol treatment.
- AEs that are related to a protocol-mandated intervention, including those that occur prior to assignment of protocol treatment (e.g. screening invasive procedures such as biopsies)

Serious Adverse Events (SAE)

A SAE is defined in general as any undesirable medical occurrence/adverse drug experience that at any dose:

- results in death (any cause, except progression of cancer under study)
- is life-threatening (i.e., the adverse event, in the view of the investigator, places the patient at immediate risk of death)
- requires or prolongs inpatient hospitalization
- results in persistent or significant disability/incapacity (i.e., the adverse event results in substantial disruption of the patient's ability to conduct normal life functions)
- is a significant medical event in the investigator's judgment
- is a congenital anomaly or birth defect (including neonatal deaths)
- is a secondary malignancy/second primary malignancy

Adverse events of special interest (AESI)

The following events are of special interest (not necessarily SAEs):

- Cases of potential drug-induced liver injury that include an elevated ALT or AST in combination with either an elevated bilirubin or clinical jaundice, as defined by Hy's Law and based on the following observations:
    - Treatment-emergent ALT or AST > 3 x ULN (or > 3 x baseline value in disease states where LFTs may be elevated at baseline) in combination with total bilirubin > 2 x ULN (of which ≥ 35% is direct bilirubin)
    - Treatment-emergent ALT or AST > 3 x ULN (or > 3 x baseline value in disease states where LFTs may be elevated at baseline) in combination with clinical jaundice
- Suspected transmission of an infectious agent by the study treatment, as defined below: Any organism, virus, or infectious particle (e.g., prion protein transmitting

transmissible spongiform encephalopathy), pathogenic or non-pathogenic, is considered an infectious agent. A transmission of an infectious agent may be suspected from clinical symptoms or laboratory findings that indicate an infection in a patient exposed to a medicinal product. This term applies only when a contamination of study treatment is suspected.

- Pneumonitis
- Colitis
- Endocrinopathies: diabetes mellitus, pancreatitis, adrenal insufficiency, hyperthyroidism, and hypophysitis
- Hepatitis, including AST or ALT > 10 x ULN
- Systemic lupus erythematosus
- Neurological disorders: Guillain-Barré syndrome, myasthenic syndrome or myasthenia gravis, and meningoencephalitis
- Events suggestive of hypersensitivity, infusion-related reactions, cytokine-release syndrome, influenza-like illness, systemic inflammatory response syndrome, and immune-mediated reactions
- Nephritis
- Ocular toxicities (e.g. uveitis, retinitis, optic neuritis)
- Myositis
- Myopathies, including rhabdomyolysis
- Grade ≥ 2 cardiac disorders (e.g. atrial fibrillation, myocarditis, pericarditis)
- Vasculitis
- Autoimmune haemolytic anaemia
- Severe cutaneous adverse reactions (SCARS) (e.g. Stevens-Johnson syndrome, dermatitis bullous, toxic epidermal necrolysis)

Severity Grade of adverse event

The adverse event severity grading scale for the CTCAE v5.0 will be used for assessing adverse event severity. For adverse events that are not specifically listed in the CTCAE, the following toxicity grading scale will be used:

Grade 1	Mild – transient or mild discomfort; no limitation in activity; no medical intervention/therapy required
Grade 2	Moderate – mild to moderate 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

The adverse event severity grade provides a qualitative assessment of the extent or intensity of a specific 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).

Causality of adverse event

Investigators should use their knowledge of the patient, the circumstances surrounding the event, and an evaluation of any potential alternative causes to determine whether an adverse event is considered to be related to protocol treatment. For patients receiving combination therapy, causality will be assessed individually for each protocol-mandated therapy.

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

Intention-to-treat (ITT) cohort: The ITT cohort of the trial includes all eligible patients randomized in the trial. Patients are evaluated in the treatment arms to which they were randomly assigned, regardless of the treatment actually received, including randomized patients who did not receive any protocol treatment.

Primary efficacy cohort: The primary efficacy cohort of the trial is the cohort based on which the primary hypothesis is formally tested. It includes all patients randomized in the trial and evaluable for the assessment of the primary endpoint (12-month PFS). That is, the primary efficacy cohort does not take into account patients lost from follow-up before a PFS event or earlier than one-year of follow-up. Patients are evaluated in the treatment arms to which they were randomly assigned, regardless of the treatment actually received, including randomized patients who did not receive any protocol treatment.

Safety cohort: The safety cohort includes all patients who have received at least one dose of the protocol treatment according to the treatment they actually received, regardless of their allocated treatment at randomization.

QoL cohort: The QoL population includes all randomized patients who have a baseline assessment and at least one on-treatment post-baseline Quality of Life Questionnaire completed.

3 Primary efficacy analysis

The **primary efficacy analysis** consists of the analysis of the primary endpoint, **12-month PFS rate**. For patients reaching 12 months after randomization without a PFS event, the first scheduled (54-week) or unscheduled tumour assessment after this timepoint, will be taken into account for the evaluation of the primary endpoint. Primary efficacy analysis will be performed separately in each treatment arm based on the primary efficacy cohort.

Formal hypothesis testing

The exact binomial one-sample test will be used for the primary efficacy hypothesis testing, within each arm separately. The conclusion will be based on examining whether or not the observed number of progression-free patients crossed the corresponding boundary (**14 or more patients among the 45 evaluable in each treatment arm should be progression-free in the 12-month timepoint in order to reject the null hypothesis** H_0: 12-month PFS rate (π_0) ≤0.18, versus the alternative hypothesis H_1: 12-month PFS rate (π_1) >0.18, evaluated at π_1=0.37). The rate of progression-free patients at 12 months (over the total number of patients evaluable at the 12-month time-point) will be accompanied by 2-side 95% exact binomial CI. In the frame of formal testing of 12-month PFS, loss from follow-up before a PFS event or earlier than one-year of follow-up will be considered as competing risks, and will not be included in the evaluable patients (primary efficacy cohort).

As previously mentioned, the trial is not designed or powered to test for statistically significant differences in the PFS between the two treatment arms. A simple selection between the two arms could be conducted based on the numerical comparison in efficacy (12-month PFS rate) along in comparison with toxicity, leading to the selection of the more promising treatment (but this will not be a statistical superiority finding).

Further PFS analyses (based on the ITT cohort)

The following PFS analyses will be also performed:

- Total number (%) of observed PFS events, 12-month PFS estimates, median PFS and respective 95% confidence intervals (CIs), by treatment arm as well as by stratification factor will be presented.
- Graphical representation of PFS via a Kaplan-Meier plot, by treatment arm and stratification factor will be performed.
- Number of PFS events, median PFS time, 12-month PFS estimates and unstratified/unadjusted HRs (along with 95% CIs), and the p-values for the comparison between the levels of each variable of interest, will be summarized by

treatment arm for the subgroups defined by the following variables of clinical interest: stratification factor (prior use of third-generation EGFR TKI [yes vs no]), sex, race, smoking status, ECOG performance status, tumor stage, EGFR mutation subtype, presence or absence of brain lesion at baseline, and age (appropriately categorized). A Kaplan Meier plot for PFS will be presented in case of significance.

- A table with information about the sites of first progression will be also provided overall and by treatment arm.
- Furthermore, to assess the effect of variables of clinicopathological interest on PFS, multivariable Cox models (stratified and unstratified; by treatment arm) will be estimated, adjusted for the clinicopathological variables of interest as defined above (in case of unstratified model, the stratification factor will be also included as covariate 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 multivariable Cox model) will be summarized in a tabular format in the report and the corresponding forest plot will be subsequently produced (if meaningful).
    - 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 and covering the full follow-up time for the majority of patients)
        - Calculation of piecewise HRs for separate time intervals.
- Also, multivariable Cox model will be estimated based on the full ITT cohort (together for both treatment arms) using the treatment arm as stratification factor.

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 the ABC-lung trial is presented.

4.1 Patient accrual, balance of stratification factor 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[1]) a table summarizing the reasons for non-randomization will be provided.
- Balance of treatment allocation by center and by stratification factor will be summarized as well.
- Patient & tumor baseline characteristics (categorical: sex, race, smoking status, ECOG performance status, tumor stage, TNM staging, EGFR mutation subtype, presence or absence of brain lesion at baseline and continuous: age at randomization, BMI), 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.
- Also available information on medical history and prior treatment will be summarised, by treatment arm and overall.

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 summarized in a table. A Kaplan-Meier plot, overall and by treatment arm, will be also provided for a graphical representation of the respective information.

[1] Registration in the database before randomization was not obligatory. Information thus is partially gathered and provided only for descriptive purposes.

- Treatment information will be summarized 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 each study drug separately. 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, 12-month TTF estimates and median TTF time along with the corresponding 95% CIs and the reasons for treatment failure will be presented overall and by treatment arm. Analogous information for TTD separately for each study drug will be presented along with the reasons of drug discontinuation (for each drug and treatment arm separately)
    - A Kaplan Meier plot for TTF as well as for TTD separately for each study drug, by treatment group, will be created.
    - For patients who progressed, information for further lines of treatment will be also provided.

4.3 Secondary efficacy analysis

Clinical efficacy (including intra-cranial/extra-cranial PFS, OS and ORR) will be further assessed separately for the two treatment arms, and by stratification factor. The secondary efficacy analysis will be based on the ITT cohort.

4.3.1 Overall survival

Similar to PFS the following will be presented for OS:

- Total number (%) of observed deaths, 12-month OS estimates, median OS and respective 95% CIs, by treatment arm as well as by stratification factor will be presented.
- Graphical representation of OS, by treatment arm and stratification factor will be performed via Kaplan-Meier plots.
- Subgroup analysis (number of deaths, median OS, 12-month OS estimates and unstratified/unadjusted HRs (along with 95% CIs), p-values) by clinicopathological variables of interest (as defined above for PFS).

- Multivariable Cox proportional hazards model (stratified and unstratified; by treatment arm), adjusted for the stratification factor and the variables of clinicopathological interest; HRs and corresponding 95% CIs for all significant OS predictors.
- Multivariable Cox model based on the full ITT cohort (together for both treatment arms) using the treatment arm as stratification factor.
- A table with information about death causes will be also provided overall and by treatment arm.

4.3.2 Intra-cranial & Extra-cranial PFS

- Summary survival tables will be presented for the secondary endpoints of intra-cranial & extra-cranial PFS.
- Furthermore, since extra-cranial progression is competing risk to intra-cranial PFS and vice versa, univariable and multivariable Fine-Grey competing risk regression will be used to assess the cumulative incidence rate of intra-cranial recurrence rate.
- Cumulative incidence plots of CNS progression, non-CNS progression, and death, will be also presented by arm.

4.3.3 Objective response rate

- Best overall responses (BOR) as well as objective response rate (ORR) will be presented overall and separately for the two treatment arms, along with a 95% exact binomial CI (for the period up to end of trial treatment as well as to end of follow-up).
- Univariate and multivariable logistic regression models of the "response" status will be further applied, adjusting for stratification factor and variables of clinical interest (by treatment arm).
- 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.

4.3.4 Duration of response

- Median DoR, along with the corresponding 95% CIs will be presented, for all responders and separately for the two treatment groups.

- Graphical representation of DoR will be also performed via Kaplan-Meier and swimmer plots, separately for each treatment arm.

4.3.5 Quality of life (QoL)

- Changes in EORTC QLQ-C30 and LC13 functional, symptom and global scales from baseline will be analysed descriptively (mean (95% CI), median, min, max) at each timepoint of interest and presented via boxplots (by treatment arm).
- Minimally important difference (MID) is defined as the smallest change in a QoL score considered important to patients. Proportion of responding patients reporting clinically meaningful change (≥ 10-point change) in EORTC QLQ-C30 and LC13 functional, symptom and global scales at different timepoints of interest will be presented with a stacked bar plot (by treatment arm).
- Furthermore, a line plot shows the mean change from baseline along with the corresponding 95% CI (error bars) over time of assessment will be produced.
- Reasons for missing data will be presented in frequency tables for each scheduled assessment without available QoL data.
- TTDet endpoint will be analysed as a time-to-event endpoint and presented separately by treatment arm. Kaplan-Meier plot will be produced.
- Longitudinal models will be also used to test changes over time in the EORTC QLQ-C30 and LC13 subscales/symptoms. Models will be adjusted for patient and disease characteristics (including age, sex, smoking status, ECOG performance status, tumor stage) and account for missing responses.

4.3.6 Subgroup analysis

Subgroup analyses of primary and secondary efficacy measures (PFS & OS) by treatment arm, will be performed as described in sections 3 and 4.3.1 according to the following:

- prior use of third-generation EGFR TKI (e.g., osimertinib) or not
- EGFR mutation subtype (del19 versus L858R)
- PD-L1 expression levels (<1%, 1-49%, ≥50% of PD-L1 expression in tumour tissue). Subgroup analysis for PD-L1 will be performed, when data are available.
- Other variables of clinical interest: sex, race, smoking status, ECOG performance status, tumor stage, presence or absence of brain lesion at baseline, and age (appropriately categorized).

Additional, analogous exploratory analyses will be performed, based on the availability of translational data (e.g., tumour mutational burden).

Notes:

1. *If there are too few events available for a meaningful analysis of a particular subgroup comparison (i.e., less than 10 events within a subgroup category), the relationship between that subgroup and the efficacy endpoint will not be formally analysed. In this case, only descriptive summaries will be provided.*
2. *No adjustment to the significance level for testing of the subgroup analyses will be made (nominal significance level for all comparisons: 5%), since all these analyses will be considered supportive of the analysis of efficacy endpoints.*

4.4 Sensitivity efficacy analysis

In a sensitivity analysis framework, the efficacy analysis will be repeated using the safety cohort. For the primary endpoint of PFS and the secondary efficacy endpoints (extra-cranial PFS, intracranial PFS and OS), the total number (%) of observed events, 12-month estimates, median times and respective 95% CIs, by treatment arm and overall will be presented.

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). The worst (AE) grade (highest toxicity) observed over the whole treatment period will be taken into account for each patient and adverse event severity will be graded according to CTCAE v5.0.

Safety analysis will include the following:

- Overview of the number of patients who experienced an AE of any grade, of grade ≥3, leading to treatment discontinuation or death, or experienced any SAE or AESI, 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 by treatment arm and overall. Analogous information will be provided for treatment-related AEs. Also, number of patients that entered the study with baseline symptoms will be reported.
- Number of events (AEs/SAEs/AESIs) and rate of occurrence per month of follow-up, by treatment arm and overall.
- Number of patients experiencing a specific number of events (AEs/SAEs/AESIs), by treatment arm and overall.

- Distribution of AEs by grade and CTCAE category, 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 only for the treatment related AEs).
- Most frequent AEs (of any grade) occurring in more than 10% overall (or any other relevant %) will be presented separately for the two treatments (analogous table only for the treatment related AEs).
- Maximum severity of AEs per patient, by treatment arm and overall.
- Number of SAEs by center.
- For all fatal SAEs, cause of death will be provided, by treatment arm and overall.

5 Technical details

Data will be primarily analysed using the SAS software package (version 9.4 or higher), 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 45 evaluable patients (per arm) reach one year of follow-up time into the trial from randomization or have experienced a PFS event (progression and/or death) earlier. 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 a 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.
- 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 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 counted only once (with the highest grade) for the presentation 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
    IIIa. Analysis of primary endpoint
    IIIb. Analysis of secondary endpoints
IV. Safety analysis
V. Sensitivity analysis

All tables and figures will be included in an appendix.