

**ETOP IBCSG Partners
13-18 BEAT-meso**

A multicentre randomised phase III trial comparing atezolizumab plus bevacizumab and standard chemotherapy versus bevacizumab and standard chemotherapy as first-line treatment for advanced malignant pleural mesothelioma

**Statistical Analysis Plan (SAP) for
Final Analysis**

A clinical trial of ETOP IBCSG Partners Foundation

Protocol version 3.1 (20210118)

SAP VERSION	V1.1
SAP DATE	20 December 2022
TRIAL STATISTICIAN	Urania Dafni
TRIAL PRIMARY INVESTIGATOR	Enriqueta Felip, Sanjay Popat
FINAL ANALYSIS SAP AUTHOR	Urania Dafni, Zoi Tsourti, Panagiota Zygoura

Contents

INTRODUCTION.....	4
1 Trial oversight	5
1.1 Previous protocol versions.....	5
1.2 Objectives.....	6
1.3 Endpoints	6
1.4 Most important eligibility criteria	7
1.5 Trial treatment	8
1.6 Statistical design, sample size & power	8
1.7 Total trial duration	9
2 Statistical considerations for final analysis	10
2.1 Analysis timing	10
2.2 Study's endpoints.....	10
2.2.1 Primary endpoint	10
2.2.2 Secondary endpoints.....	10
2.2.3 Exploratory endpoints.....	12
2.3 (Serious) Adverse Events.....	12
2.4 Analysis populations	14
3 Primary efficacy analysis of overall survival (OS).....	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	17
4.3 Secondary and exploratory endpoints.....	17
4.3.1 Progression-free survival.....	17
4.3.2 Objective response rate & disease control rate.....	18
4.3.3 Duration of response.....	18
4.3.4 Quality of life (QoL)	19
4.3.5 Subgroup analysis.....	19

4.4	Sensitivity efficacy analysis	20
4.5	Safety analysis.....	21
5	Technical details.....	22
5.1	Data Retrieval Information	22
5.2	Missing Data.....	22
5.3	Reporting conventions.....	23
5.4	Multiple recordings of an event for the same patient.....	23
5.5	Presentation of results.....	23

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 BEAT-meso trial to be implemented (based on protocol version 3.1 (20210118)).

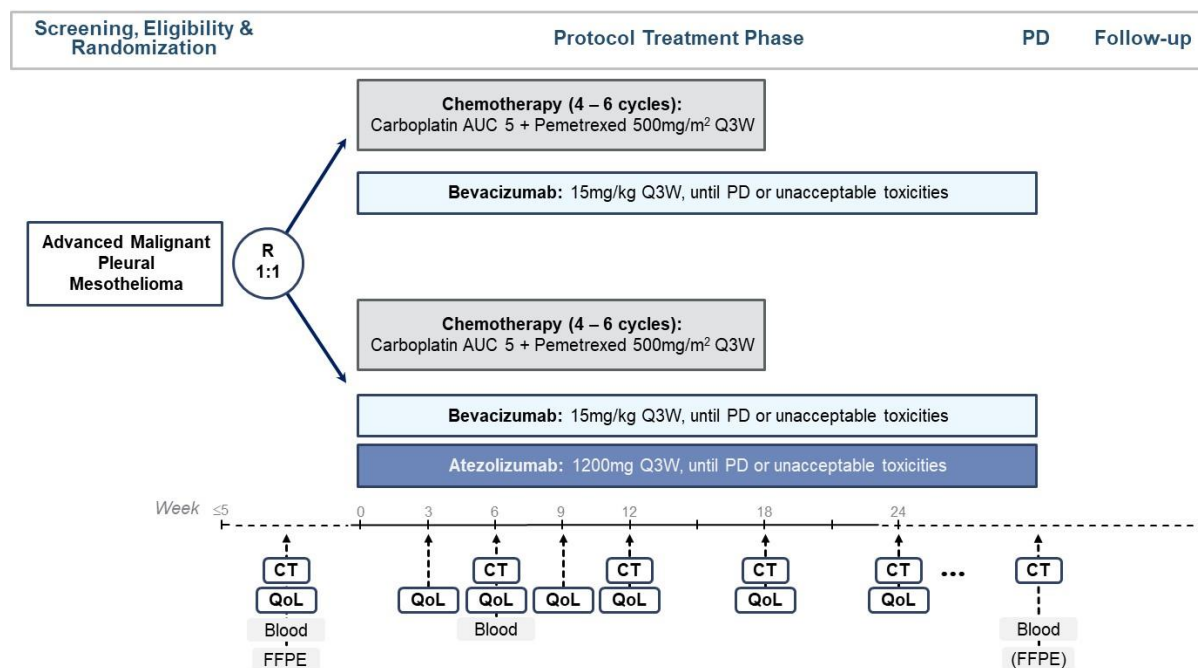
A short description of the contents of this statistical analysis plan is provided below:

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 overall survival (OS)**
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**

Important note: In case a statistically significant OS benefit is identified at the scheduled interim analysis for efficacy (i.e., crossing the pre-specified boundary), a complete analysis for OS (and the rest of endpoints) will be performed according to this SAP.

1 Trial oversight

This is a two-arm, stratified, randomised, open-label multicentre phase III trial, with an interim efficacy analysis, evaluating the activity of atezolizumab when added to standard of care (carboplatin/pemetrexed/bevacizumab), as first-line treatment for advanced malignant pleural mesothelioma.



SCHEMA 1. Trial design

Target Population: Histologically-confirmed advanced malignant pleural mesothelioma (MPM), not amenable for surgery based on local standards (all subtypes are eligible).

BEAT-meso is a trial with block stratified randomization balanced by institution and patients randomized 1:1 to the experimental and control arms. The two stratifications factors are:

- Histological subtype (pure epithelioid versus not)
- Stage (IV according to 8th TNM classification versus others)

1.1 Previous protocol versions

The BEAT-meso trial has been actively recruiting patients under the original protocol since April 30th, 2019. Two protocol amendments have been approved so far.

The main changes that occurred in the latest protocol version 3.1 (20210118) based on amendment 2 have been the following:

- The statistical design was modified so as to assess overall survival as the primary endpoint of the study (instead of co-primary OS and PFS). The original target median OS for the two arms and the corresponding target HR remained unchanged (median OS from 17 to 24 months, HR 0.708), while the total OS events, sample size, significance level and duration of the study were appropriately adjusted. In amendment 2, PFS will be tested as part of the secondary endpoints.
- A single interim efficacy analysis is planned to be conducted for the primary endpoint of OS, at 75% of information fraction (instead of the originally planned 50%).

1.2 Objectives

Primary objective

The **primary objective** of the study is to assess the efficacy of atezolizumab in terms of overall survival when added to standard of care (carboplatin/pemetrexed/bevacizumab), as first-line treatment of advanced MPM.

Secondary objectives

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

- To evaluate secondary measures of clinical efficacy including investigator assessed PFS, objective response rate, disease control rate, time to treatment failure, duration of response.
- To assess the safety and tolerability of the treatment.
- To evaluate symptom-specific and global quality of life (QoL).

1.3 Endpoints

Primary endpoint:

- Overall survival (OS)

Secondary endpoints:

- Progression-free survival (investigator assessed) according to the mRECIST v1.1
- Objective response rate (ORR) defined as the percentage of patients that achieve CR or PR according to the mRECIST v1.1
- Disease control rate (DCR) defined as the percentage of patients with complete or partial response, or disease stabilization at 24 weeks

- Time-to-Treatment Failure (TTF) defined as the time from the date of randomisation to discontinuation of protocol treatment for any reason
- Duration of Response (DoR) defined as the time from the date of first documented objective response to the date of first documented progression or death
- Adverse events graded according to CTCAE v5.0
- Symptom-specific and global QoL assessed by the mesothelioma version of the Lung Cancer Symptom Scale (LCSS-Meso)

1.4 Most important eligibility criteria

Inclusion criteria at randomization:

- Histologically confirmed advanced malignant pleural mesothelioma (all histological subtypes are eligible)
- Able to understand and give written informed consent and comply with trial procedures
- Age ≥ 18 years
- Performance Status 0-1
- Not amenable for radical surgery based on local standards
- Availability of tumor tissue for translational research
- Evaluable disease or measurable disease as assessed according to the modified RECIST v1.1
- Life expectancy ≥ 3 months
- Adequate haematological, renal (CrCl ≥ 45) and liver function
- Completed baseline QoL questionnaire
- Men and women of childbearing potential must agree to use adequate contraception

Exclusion criteria at randomization:

- Prior treatment for malignant pleural mesothelioma

- Active autoimmune disease that has required systemic treatment in the past 2 years
- Previous history of significant haemoptysis (defined as at least 2.5mL emission of red blood) within 3 months prior to inclusion
- Recent surgery:
 - Major surgery or significant traumatic injury within 28 days prior to the first dose of bevacizumab
 - Minor surgical procedure within 7 days, or placement of a vascular access device within 2 days prior to the first dose of bevacizumab
- HIV or active hepatitis B or hepatitis C

1.5 Trial treatment

Experimental arm (ABC):

- 4-6 cycles chemotherapy:
Carboplatin AUC 5 + pemetrexed 500 mg/m², D1Q3W plus
- Bevacizumab 15mg/kg D1Q3W until disease progression, refusal or unacceptable toxicity plus
- Atezolizumab 1200 mg D1Q3W until disease progression, refusal or unacceptable toxicity

Control arm (BC):

- 4-6 cycles chemotherapy:
Carboplatin AUC 5 + pemetrexed 500 mg/m², D1Q3W plus
- Bevacizumab 15mg/kg D1Q3W until disease progression, refusal or unacceptable toxicity

1.6 Statistical design, sample size & power

BEAT-meso is a 1:1 randomized phase III trial, with OS as primary endpoint, taken into account in the statistical design.

Comparison of OS between the two arms

The **targeted median OS** for patients treated with atezolizumab plus bevacizumab and chemotherapy is **24 months**, which corresponds to an absolute **increase of 7 months compared to a median OS of 17 months** for patients treated with bevacizumab and chemotherapy alone (corresponding to an HR of 0.708).

Using the log-rank test at a **one-sided significance level of 2.5%**, a total of **284 OS events** are required in order to detect with **82.1% power** the targeted **increase of 7 months in the median OS**.

Assuming an accrual rate of 8 patients per month for the first 15 months and 20 patients per month thereafter, and a 5% loss to follow-up by month 24, a total of **400 randomised patients**, 200 in each treatment group, are needed to be followed for an expected duration of 58 months so as to observe the required number of 284 OS events.

Interim efficacy analysis (summary)

One interim analysis is planned for this study. The planned interim efficacy analysis for OS will be performed when approximately **213 OS events (75% of the total planned events)** have been observed in the efficacy cohort. This is expected to occur approximately 42 months after the randomisation of the first patient (power 57%). If the efficacy boundary for the primary endpoint OS is crossed at the interim analysis, superiority of the experimental treatment with respect to OS will be claimed and this will be considered the primary analysis. If the interim boundary is not crossed at the interim analysis, the study will be continued for the final analysis to derive conclusions for OS. The interim analysis statistical significance is determined based on the Lan DeMets spending function with an O'Brien-Fleming boundary. In case that a slightly larger number of OS events is used, the corresponding boundary will be revised as needed based on the actual number of events observed at the interim (according, as before, to the Lan-DeMets spending function).

Interim efficacy analysis and corresponding testing strategy are described in detail in the interim analysis SAP.

1.7 Total trial duration

The total trial follow-up duration to observe the required events is expected to be 58 months after randomisation of the first patient with an interim efficacy analysis based on O'Brien-

Fleming boundary at approximately 42 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 6 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 either after the interim analysis, if the efficacy boundary for OS is crossed in favour of the alternative, or (if the efficacy boundary for OS is not crossed in favour of the alternative) when the 284 required OS events will be available for the BEAT-meso randomized patients.

2.2 Study's endpoints

2.2.1 Primary endpoint

The primary endpoint of the trial is OS, defined as the time from the date of randomization until death due to any cause. Data for patients without documented death at the time of the final analysis will be censored at the date when they were last known to be alive (last follow-up). Data for patients without post-baseline information will be censored at the date of randomization (plus 1 day).

2.2.2 Secondary endpoints

Secondary endpoints, according to the protocol, include PFS, ORR, disease control, TTF, DoR adverse events (AEs) and QoL. More specifically:

- PFS is defined as the time from the date of randomisation to the first documented disease progression (investigator assessed according to the mRECIST v1.1) or death, if progression is not documented. Censoring (for patients without progression/death) will occur at the date of last tumor assessment. As 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.

Patients without any post-baseline tumour assessment will be censored at the date of randomization (plus 1 day).

- ORR is defined as the proportion of patients that achieve a best overall response (complete response (CR) or partial response (PR)) across all post-randomization timepoints, until the end of follow-up for progression. Alternatively, all post-randomization time-points, until termination of all trial treatments will be taken into account for the determination of best overall response. Responses will be investigator assessed according to the mRECIST 1.1.
- Disease control is defined as complete or partial response, or disease stabilization, at 24 weeks. Confirmation of response will not be required.
- TTF is defined as the time from the date of randomisation to discontinuation of at least one of the drugs consisting the protocol treatment due to any reason, such as toxicity, investigator decision or refusal, or discontinuation due to other reasons (including progression, death or withdrawal/lost to follow-up (LFU)). Censoring for TTF (patients on treatment) will occur at the last follow-up date.
- DoR is defined as the time interval from the date of first documented objective response (CR or PR, according to the mRECIST v1.1) to the date of first documented progression or death. Censoring will occur at the last tumor assessment with response other than progression.
- Toxicity, defined as AEs (any-cause as well as treatment-related) graded according to CTCAE v5.0.
- Symptom-specific and global QoL, assessed by the LCSS-Meso, an 8-item questionnaire including five symptom items associated with mesothelioma (anorexia/loss of appetite, fatigue, cough, dyspnea and pain) and three global items for overall symptom burden and HRQoL (symptomatic distress, interference with activity level and global QoL). Each item is scored from 0 to 100, with 0 representing no symptom distress, no interference with activity level, or best possible health-related quality of life. Change of ≥ 10 points from baseline will be considered as the minimally important difference (MID) based on prior research. The primary QoL endpoint will be the change in the LCSS total score (average of all 8 items) from baseline to 12 weeks after treatment start.

2.2.3 Exploratory endpoints

- Similar to TTF, time to treatment discontinuation (TTD) for each drug separately (atezolizumab, bevacizumab or chemotherapy) is defined as the time from the date of randomization until discontinuation of the specific protocol drug for any reason.

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 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 signature of informed 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 IBCSG Partners Foundation 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.

QoL cohort: The QoL population includes all randomized patients who had a baseline assessment and at least one on-treatment post-baseline score for any LCSS-Meso item.

3 Primary efficacy analysis of overall survival (OS)

Primary efficacy analysis of OS (primary endpoint) will be performed on all randomized patients, based on their initial treatment assignment (ITT, efficacy cohort), stratified by the randomisation stratification factors (histological subtype and stage). OS time is measured from randomization.

Formal hypothesis testing

The study is designed to test the hypothesis that addition of atezolizumab to bevacizumab and chemotherapy will lead to an increase in median OS to 24 months, from 17 months with bevacizumab and chemotherapy alone. **According to the study design, this corresponds to a HR of 0.708.** Using the log-rank test at a **one-sided significance level of 2.5%**, a total of **284 deaths** are required in order to achieve the trial goal with **82.1% power**, while an **interim analysis** is scheduled to be performed at **75% of the information time**, i.e. when approximately **213 deaths** have been observed. In order to maintain the overall one-sided type I error at 2.5% according to the O'Brien-Fleming approach (calculated by the Lan-DeMets spending function), the one-sided p-value boundary for the scheduled interim is 0.0096, while the exact **p-value boundary for the final efficacy analysis is 0.0221**. In case that a slightly different number of OS events (>284 according to the design) is used for the final analysis, the corresponding p-value boundary will be recalculated based on the same methodology and the exact number of observed deaths in the ITT cohort.

In the frame of final efficacy analysis, the **formal comparison of the OS between the two treatment arms**, will be based on the **stratified one-sided log-rank test** (with histologic subtype and stage 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 OS analyses

The following OS analyses will be also performed and presented

- Total number (%) of deaths due to any cause, 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
- Number of deaths, median OS and unstratified/unadjusted HRs (along with 95% CIs), interaction p-value between treatment and each variable of interest, will be summarized for the subgroups defined by treatment and the following variables of interest: stratification factors (histological subtype [pure epithelioid vs not] and stage [IV vs other]), sex, age (appropriately categorized), smoking status, ECOG performance status at diagnosis, EORTC prognostic score for malignant mesothelioma, PD-L1 expression levels (appropriately categorized) and medical history. This information will be depicted in a tabular format in the report and a forest plot will be produced for the publication/presentation.
- Summary table of the death causes (e.g., lung cancer, toxicity, other, etc) overall and by treatment arm.
- Furthermore, to assess the effect of trial treatment and other clinicopathological variables on OS, Cox proportional hazards model will be fitted.
 - o Initially a univariate (stratified and unstratified) Cox model will be fitted.
 - o Subsequently, multivariable 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 multivariable Cox model) will be summarized 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 and covering the full follow-up time for the majority of patients)
 - Calculation of piecewise HRs 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 the BEAT-meso 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 summarized as well.
- Patient & tumor baseline characteristics (categorical: ethnicity, sex, smoking status, ECOG performance status at diagnosis, EORTC prognostic score, tumor stage, histological subtype, PD-L1 expression level, TNM stage, 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 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 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.

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 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, 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 and exploratory endpoints

4.3.1 Progression-free survival

Similar to OS the following will be presented for PFS:

- The total number (%) of PFS events, overall and by treatment arm (as well as by stratification factor) will be presented.
- 1/2-year PFS estimates, median PFS and respective 95% CIs will be provided (comparison between the arms based on stratified and unstratified log-rank test)
- Graphical representation of PFS, by treatment arm and separately for each stratum will be performed via Kaplan-Meier plots.
- Subgroup analysis (number of PFS events, median PFS and unstratified/unadjusted HRs (along with 95% CIs)) by treatment and clinicopathological variables of interest (as defined above for OS).
- Univariate and multivariable Cox proportional hazards model, adjusted for the stratification factors and the clinicopathological variables of interest; HRs and corresponding 95% CIs for all significant PFS predictors
- A table with information about the sites of progression will be also provided.

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 (for the period up to end of follow-up as well as to end of trial treatment as specified in section 2.2.2).
- 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

- Median DoR, along with the corresponding 95% CIs will be presented, for all responders and separately for the two treatment groups.
- Duration of response will be compared between the two treatment arms using the Kaplan-Meier method and the stratified log-rank test.

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

4.3.4 Quality of life (QoL)

The QoL of patients will be repeatedly measured by the LCSS-Meso scale at baseline before randomisation, and within 3 days before treatment administration in cycles 2-5 (e.g., at weeks 3, 6, 9 and 12) and cycles 7 and 9 (e.g., at weeks 18 and 24) OR until end of protocol treatment, whatever is first. The QoL analysis will be performed on the QoL cohort.

The primary QoL analysis is to compare the change in the LCSS-Meso total score (average of all 8 items) from baseline to 12 weeks. Changes will be assessed using a Wilcoxon matched pairs test. A sub-score using the mean of all five major symptoms or “average symptom burden index” (ASBI), the single global QoL item, and individual items to report specific areas of change will be presented as secondary descriptive measures of QoL. These will be summarized with mean (SD) and medians (range) and will be graphically depicted by longitudinal plots at each post-baseline time point.

In addition, mixed-effects linear regression modelling for repeated measures will be used to test the effect of treatment on changes in the LCSS-Meso total score, the ASBI, and each single item (i.e., 5 symptom items, 3 global items) with an unstructured covariance structure. The model will include the treatment assignment (chemotherapy plus bevacizumab versus chemotherapy plus bevacizumab plus atezolizumab), assessment time-point (categorical: 3, 6, 9, 12, 18, and 24 weeks), and the interactions of the two variables. Models will be adjusted for patient and disease characteristics (including age, sex, smoking status, ECOG performance status, stage) and account for missing responses. From the model, an estimated least squares mean difference between treatment groups will be calculated along with the corresponding 95% CI and compared at 12 and 24 weeks using model contrasts.

Reasons for missing data will be assessed for each scheduled assessment with no available QoL data and presented in frequency tables by treatment arm.

4.3.5 Subgroup analysis

The purpose of the subgroup analyses is to determine whether the treatment effect is consistent across expected prognostic and/or predictive factors. The between-group treatment effect for all efficacy endpoints will be estimated within each category of the following pre-specified variables.

Main subgroup analyses:

- Histological subtype
- Stage
- PD-L1

Other pre-planned subgroup analyses:

- Age group
- Mesothelioma risk score (MRS) (3 levels: 0, 1 or 2 risk factors) with the risk factors being baseline systemic immune-inflammatory index (SII) (good: <1,250 vs poor: ≥1,250) and baseline haemoglobin (good: ≥LLN vs poor: <LLN) (LLN = 135 for males, LLN = 120 for females)¹.
- Tumor Mutation Burden at baseline (if available)

Notes:

1. For each subgroup level, the HR (for the treatment comparisons of interest) and 95% CI will be calculated from a Cox proportional hazards model that only contains a term for treatment. These will be presented on a forest plot including also the results of the overall primary analysis.
2. The interaction of each subgroup with the treatment effect on OS and PFS will be tested based on respective Cox models (and it will be considered statistically significant at 10% 2-sided significance level).
3. 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.
4. 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.

¹ Banna et al. Lung Cancer. 2022 Jul;169:77-83.

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 summarized 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. $\geq 10\%$) 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 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 total number of 284 deaths required according to the statistical design of the trial are observed (or on the database used for the interim analysis, if the efficacy boundary for OS is crossed in favour of the alternative at the interim analysis). 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 (eg, 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.
- 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 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

All tables and figures will be included in an appendix.