Oncology Statistical Analysis Plan
Study Code D5169C0001
Edition Number 2.0
Date 09 Jan 2023

Oncology Statistical Analysis Plan

Study Code	D5169C00001
Edition Number	2.0
Date	09 Jan 2023

A Phase III, Open-label, Randomized Study of Osimertinib with or without Platinum Plus Pemetrexed Chemotherapy, as First-line Treatment in Patients with Epidermal Gr wth Factor Receptor (EGFR) Mutation-Positive, Locally Advanced or Metastatic Non-small Cell Lung Cancer (FLAURA2)

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE(s)	Adverse event (s)
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ANCOVA	Analysis of Covariance
AST	Aspartate transaminase
ATC	Anatomical Therapeutic Chemical
AUC5	AUC of 5 mg/mL/min
AUC_{ss}	Area under plasma concentration-time curve during any dosing interval at steady state [amount*time/volume]
BDRM	Blinded data review meeting
BICR	Blinded Independent Central Review
B	Blood
BoR	Best objective response
BP	Blood pressure
cFAS	CNS Full Analysis Set
CI	Confidence interval
CLIA	Clinical Laboratory Improvement Amendments
CL_{ss}/F	Apparent otal body clearance at steady state
cMET	Hepatocyte Growth Factor Receptor
$C_{max,ss}$	Maximum plasma concentration at steady state
$C_{min,ss}$	Minimum plasma concentration at steady state
CMH	Cochran–Mantel–Haenszel
CNS	Central nervous system
CR	Complete response
CRF	Case report form
CRO	Contract research organization
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed tomography
CTx	Chemotherapy

Abbreviation or special term	Explanation
CTCAE	Common Terminology Criteria for Adverse Event
ctDNA	Circulating tumor deoxyribonucleic acid
CV	Coefficient of variation
DAE	Discontinuation of investigational product due to adverse events
DBL	Database lock
DCO	Data cut-off
DCR	Disease control rate
DoR	Duration of response
d.p.	Decimal place
eCRF	Electronic case report form
ECG	Electrocardiogram
EGFR	Epidermal growth factor receptor
EGFRm+	Epidermal growth factor receptor mutation-positive
EORTC	European Organization f r Research and Treatment of Cancer
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer - Quality of Life Questionnair – Core 30 items
EORTC QLQ-LC13	European Organiz tion for Research and Treatment of Cancer - Quality of Life Questionnaire – Lung Cancer module
EQ-5D-5L	EuroQoL five dimensions, five level
Ex19del	Exon 19 del tions
FAS	Full Analysis Set
GGT	Gamma-Glutamyl Transpeptidase
Hb	Hemoglobin
HER	Human Epidermal Growth Factor Receptor
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hazard ratio
HRQoL	Health related quality of life
ICU	Intensive care unit
IDMC	Independent Data Monitoring Committee
IF	Information Fraction
ILD	Interstitial Lung Disease
IP	Investigational product
ITT	Intention-to-treat

Abbreviation or special term	Explanation
K-M	Kaplan-Meier
L858R	Exon 21
LD	Longest diameter
LDH	Lactate dehydrogenase
LRCI	Likelihood ratio confidence interval
LLoQ	Lower level of quantification
LSI	Last subject in
lsmean	Least squares mean
LVEF	Left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
MET	Tyrosine-protein kinase Met
MMRM	Mixed Model for Repeated Measures
MRI	Magnetic resonance imaging
MUGA	Multi Gated Acquisition Scan
NA	Not applicable
NC	Non-calculable
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NL	New le ion
NSCLC	Non-small cell lung cancer
NTL	Non-target lesion
OAE	Other significant adverse events
ORR	Objective response rate
OS	Overall survival
OS12	Proportion of patients alive at 12 months
PAP	Payer Analysis Plan
PD	Progressive disease
PGIS	Patients Global Impression of Severity
PFS	Progression free survival
PFS2	Time from randomization to second progression on a subsequent treatment
PID	Percentage intended dose
PK	Pharmacokinetics

Abbreviation or special term	Explanation
PR	Partial response
PRO	Patient reported outcome
PRO-CTCAE	Patient Reported Outcome version of the Common Terminology Criteria for Adverse Event
PS	Performance status
PT	Preferred term
PSI	Pulmonary symptom index
Q21d	Every 21 days
Q3W	Every 3 weeks
QD	Once daily
QoL	Quality of life
QT	Interval on the electrocardiogram representing the duration of depolarization and repolarization of the heart
QTc	Corrected QT interval
QTcF	Fredericia's corrected QT interval
RBC	Red Blood Cell
RDI	Relative dose intensity
RECIST	Response Eva uation Criteria in Solid Tumours
RECIST 1.1	Response Evaluation Criteria in Solid Tumours version 1.1
REML	Restricted maximum likelihood
S	Serum
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAF	Safety analysis set
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SD	Stable disease
SMQ	Standardized MedDRA query
SoA	Schedule of assessments
SOC	System organ class
SRC	Safety review committee
STx	Subsequent treatment
TBIL	Total bilirubin
TFST	Time to first subsequent therapy or death

Abbreviation or special term	Explanation
TKI	Tyrosine kinase inhibitor
TL	Target lesion
$T_{max, ss}$	Time to $C_{max, ss}$
TP53	Tumor protein 53
TSST	Time to second subsequent treatment
TTD	Time to deterioration
U	Urine
ULN	Upper limit of normal
ULoQ	Upper level of quantification
US	United States
WHO	World Health Organization

1. STUDY DETAILS

This statistical analysis plan (SAP) contains a more detailed description of the analyses summarized in the FLAURA2 clinical study protocol (CSP). This SAP is based on version 2.0 of the CSP (26 August 2021).

1.1 Study Objectives

1.1.1 Study objectives for safety run-in period

Prior to the start of the randomized period of the study, a non-randomized safety run-in will allocate approximately 30 patients to evaluate the safety and tolerability of osimertinib with platinum chemotherapy (carboplatin or cisplatin) followed by maintenance pemetrexed (approximately 15 patients per choice of platinum chemotherapy).

Primary Objective:	**Endpoint/Variable:**
To evaluate the safety and tolerability of osimertinib plus chemotherapy	Adverse events graded by CTCAE v5; Clinical ch mis y, hematology and urinalysis; Vital signs (pulse and blood pressure); physical examination; weight; LVEF; ECG parameters; WHO Performance Status
Secondary Objective:	**Endpoint/Variable:**
To assess the efficacy of osimertinib plus chemotherapy	ORR, DoR; depth of response; DCR by Investigator; OS;
To assess the PK of osimertinib when given with chemotherapy	Steady-state plasma concentrations and appropriate PK parameters (CL_{ss}/F, $C_{max,ss}$ $C_{min,ss}$ and AUC_{ss}) of osimertinib and its metabolite, AZ5104 will be summarized. *
Exploratory	**Endpoint/Variable:**
To explore how changes in plasma-based biomarkers (eg, ctDNA, proteomic) correlate with response	Quantitative ctDNA analysis using specific EGFR biomarkers or broader cancer biomarker panel in longitudinal plasma samples, to assess ctDNA clearance and correlate with response (eg, PFS)
To collect and store DNA for genetic research to explore how genetic variations may affect clinical parameters, risk and prognosis of diseases, and the response to medications	Correlation of polymorphisms with variation in PK, pharmacodynamics, safety or response observed in patients treated with osimertinib plus chemotherapy
To explore efficacy biomarkers and biomarker changes in baseline, longitudinal and progression samples (plasma and tumour tissue) for correlation with response	Assessment of innate and acquired resistance mechanisms and biomarkers of response including but not limited to mutations in, amplifications and expression of EGFR, TP53, HER2, MET and relevant pathway genes Proteomic and/or gene expression analysis eg, biomarkers of inflammation

Exploratory	Endpoint/Variable:
To collect and store tumour, serum and plasma samples for potential exploratory research into factors that may influence susceptibility to development of NSCLC and/or response to osimertinib and/or chemotherapy (where response is defined broadly to include efficacy, tolerability or safety) and to assess the relationship between tissue and/or bloodborne biomarkers and selected efficacy endpoints. Tissue and plasma samples may be used to support diagnostic development.	Key genetic, gene expression and proteomic markers to include, but not limited to, EGFR mutations, HER, and proto-oncogene encoding cMET expression and/or amplification. Relationship between PK and blood-borne biomarkers. Diagnostic development.

* If feasible, further PK parameters may be derived using population PK analysis and reported separately from the CSR. Data from this study may also form part of a pooled analysis with data from other studies

1.1.2 Study objectives for the randomized period

Following a positive decision based on recommendation by the Safety Review Committee (SRC) following evaluation of data from the safety run-in period, the randomized period will begin to randomize approximately 556 new patients in a 1:1 ratio to receive osimertinib or osimertinib in combination with chemotherapy (either cisplatin or carboplatin followed by maintenance pemetrexed; investigator's choice).

Primary Objective:	Endpoint/Variable:
To assess the efficacy of osimertinib plus chemotherapy treatment compared with osimertinib	PFS using Investigator assessment as defined by RECIST 1.1; Sensitivity analysis of PFS using BICR assessment as defined by RECIST 1.1
Secondary Objective:	**Endpoint/Variable:**
To further assess the efficacy of osimertinib plus chemotherapy compared with osimertinib	OS; Landmark OS at 1, 2, and 3 years; ORR, DoR; depth of response; DCR by Investigator
To further assess the efficacy of osimertinib plus chemotherapy compared with osimertinib post progression	PFS2; TFST; TSST
To assess disease-related symptoms and health-related QoL in patients treated with osimertinib plus chemotherapy compared with osimertinib	Change from baseline and time to deterioration in EORTC QLQ-C30; Change from baseline and time to deterioration in EORTC QLQ-LC13
To assess the PK of osimertinib when given with or without chemotherapy	Steady-state plasma concentrations and appropriate PK parameters (CL_{ss}/F, $C_{max,ss}$ $C_{min,ss}$ and AUC_{ss}) of osimertinib and its metabolite, AZ5104 will be summarized. *
To compare the local EGFR mutation test result used for patient selection with the retrospective central cobas® EGFR Mutation Test v2 results from baseline tumor samples	Concordance of EGFR mutation status between the local and central cobas® EGFR Mutation Test v2 results from tumor samples with evaluable results

Secondary Objective:	**Endpoint/Variable:**
To determine efficacy of osimertinib monotherapy vs. osimertinib combined with chemotherapy based on the cobas® EGFR Mutation Test v2 plasma result for Exon 19 deletions or L858R EGFR mutations	PFS by Investigator by plasma EGFR mutation status
Safety Objective:	**Endpoint/Variable:**
To evaluate the safety and tolerability of osimertinib plus chemotherapy compared with osimertinib	Adverse events graded by CTCAE v5.0; Clinical chemistry, hematology and urinalysis; Vital signs (pulse and blood pressure); physical examination; weight; LVEF; ECG parameters; WHO Performance Status
Exploratory	**Endpoint/Variable:**
To assess the impact of osimertinib plus chemotherapy compared with osimertinib on patient reported treatment related symptoms	PRO-CTCAE symptoms
To assess patients' overall impression of severity of cancer symptoms	PGIS
To compare health resource use associated with osimertinib plus chemotherapy treatment versus osimertinib	Health Resource Use Module
To assess the impact of osimertinib plus chemotherapy compared with osimertinib on patien reported health state utility	EQ-5D-5L
To assess the efficacy of osimertinib plus chemotherapy treatment compared with osimertinib on CNS metastases in patients with CNS metastases at baseline	Neuro-radiologist assessments according to CNS RECIST 1.1 to calculate: • CNS PFS; CNS ORR; CNS DoR; CNS DCR; • Best percentage change in CNS tumor size (target lesion)
To assess the efficacy of osimertinib plus chemotherapy treatment compared with osimertinib on the prevention of CNS metastases	Neuro-radiologist assessments according to CNS RECIST 1.1 to determine the presence/absence of CNS lesions at progression in patients without CNS metastases at baseline
To compare the concordance of cobas® EGFR Mutation Test v2 versus an EGFR tissue testing alternative methods for diagnostic development	Concordance of EGFR mutation status between the cobas® EGFR Mutation Test v2 and an alternative device.
To explore how changes in plasma-based biomarkers (eg, ctDNA, proteomic) correlate with response	Quantitative ctDNA analysis using specific EGFR biomarkers or broader cancer biomarker panel in longitudinal plasma samples, to assess ctDNA clearance and correlate with response (eg, PFS)

Exploratory	Endpoint/Variable:
To collect and store DNA for genetic research to explore how genetic variations may affect clinical parameters, risk and prognosis of diseases, and the response to medications	Correlation of polymorphisms with variation in PK, pharmacodynamics, safety or response observed in patients treated with osimertinib plus chemotherapy compared with osimertinib
To explore efficacy biomarkers and biomarker changes in baseline, longitudinal and progression samples (plasma and tumour tissue) for correlation with response	Assessment of innate and acquired resistance mechanisms and biomarkers of response including but not limited to mutations in, amplifications and expression of EGFR, TP53, HER2, MET and relevant pathway genes Proteomic and/or gene expression analysis eg, biomarkers of inflammation
To collect and store tumour, serum and plasma samples for potential exploratory research into factors that may influence susceptibility to development of NSCLC and/or response to osimertinib and/or chemotherapy (where response is defined broadly to include efficacy, tolerability or safety) and to assess the relationship between tissue and/or bloodborne biomarkers and selected efficacy endpoints. Tissue and plasma samples may be used to support diagnostic development.	Key genetic, gene expression and proteomic markers to include, but not limited to, EGFR mutations, HER, and proto-oncog ne encoding cMET expression and/or amplification. Relationship between PK and blood-borne biomarkers. Diagnostic development.

* If feasible, further PK parameters may be derived using population PK analysis and reported separately from the CSR. Data from this study may also form part of a pooled analysis with data from other studies

1.2 Study Design

This is a global Phase III, open-label, randomized study of osimertinib with or without platinum chemotherapy (carboplatin, cisplatin; in estigator's choice) plus pemetrexed maintenance chemotherapy conducted in patients with locally-advanced or metastatic EGFRm (Ex19del and/or L858R) non-small cell lung cancer (NSCLC) who have not received any prior therapy for advanced disease.

The proposed study will include patients with EGFRm NSCLC who have either: (1) a pre-existing positive (Ex19del or L858R) tissue test result obtained from a CLIA-certified local laboratory (for United States (US) sites) or from an accredited local laboratory (for sites outside of the US); or (2) have a positive tissue Ex19del or L858R EGFR mutation test based on the cobas® EGFR Mutation Test v2 conducted prospectively in a central laboratory.

Prior to the start of the randomized period, approximately 30 patients will be studied in a non-randomized fashion in 2 cohorts of patients to evaluate the safety and tolerability of osimertinib with platinum-based chemotherapy (carboplatin or cisplatin) followed by pemetrexed chemotherapy (Figure 1). Approximately 15 patients per cohort will receive osimertinib 80 mg QD with either cisplatin (75 mg/m^2) or carboplatin (AUC5), and pemetrexed (500 mg/m^2), both administered Q3W for 4 cycles, followed by osimertinib 80 mg QD plus pemetrexed

maintenance (500 mg/m^2) Q3W until RECIST 1.1 defined progression or another discontinuation criterion is met.

When at least 12 patients in each cohort have either received ≥ 3 cycles of study treatment (osimertinib, cisplatin or carboplatin, and pemetrexed) or have discontinued study treatment due to unacceptable toxicity, a Safety Review Committee will convene. A formal database lock (DBL) will be required to facilitate this data review. Full details of the Safety Review Committee procedures and processes can be found in the Safety Review Committee Charter. All data, including safety, tolerability, and available pharmacokinetics (PK) data from all patients, will be reviewed. The Safety Review Committee will include independent experts with relevant experience in clinical trial conduct, methodology, and procedures in patients with NSCLC, and AstraZeneca personnel. Based on these data and taking into consideration data from other sources (eg, any updated information from the Phase II study of osimertinib alone vs. osimertinib plus carboplatin/pemetrexed [TAKUMI Study LOGIK1604/NEJ032A]), the Safety Review Committee will recommend whether the data support the initiation of the randomized period of the study after which it will end its function.

Following Safety Review Committee evaluation and recommendation to continue the study into the phase 3 randomized component, patients who particip ted in the safety run-in will not be included among the patients in the randomized period but may continue their allocated treatment per protocol.

Following a positive decision by the Safety Review Committee based on the evaluation of data from the safety run-in period, the randomized period will begin to randomize approximately 556 new patients, randomized in a 1:1 ratio to receive osimertinib alone or osimertinib with pemetrexed and either cisplatin or carboplatin. Patients will be stratified by race (Chinese/Asian vs. non-Chinese/Asian vs. non-Asian), WHO PS (0 vs. 1), and method for tissue testing (central vs. local) for generation of the p-value, using the Efron approach for handling ties. It is anticipated that approximately 60% Asian patients and 40% non-Asian patients will be recruited. The Investigator will decide before randomization which chemotherapy regimen (carboplatin/pemetrexed or cisplatin/pemetrexed) a patient would receive in case the patient is assigned to the osimertinib plus chemotherapy arm. In the event that the Safety Review Committee deems one of the study treatments to be inappropriate, only one study treatment may be recommended for the randomized period (eg, osimertinib with pemetrexed and cisplatin or osimertinib with pemetrexed and carboplatin).

The two treatment regimens will be as follows:

- Osimertinib 80 mg once daily
- Osimertinib 80 mg once daily with pemetrexed (500 mg/m^2) (with vitamin supplementation) plus either cisplatin (75 mg/m^2) or carboplatin (AUC5), with both treatments administered on Day 1 of 21-day cycles for 4 cycles, followed by osimertinib 80 mg daily plus pemetrexed maintenance (500 mg/m^2) Q3W

Randomized treatment will continue until RECIST 1.1 defined progression by Investigator or until another discontinuation criterion is met.

At the Investigator's discretion, study treatment may continue for as long as a patient continues to derive clinical benefit after RECIST 1.1 progression in the absence of any discontinuation criteria. However, if the patient is deemed to have clinically significant unacceptable or irreversible toxicities, rapid tumor progression, or symptomatic progression requiring urgent medical intervention (eg, CNS metastases, respiratory failure, spinal cord compression) study treatment must be discontinued.

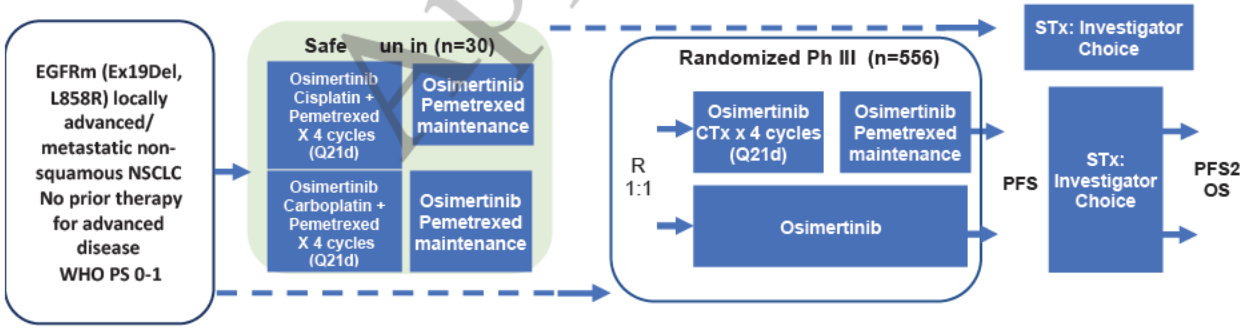
Following treatment discontinuation, subsequent therapy will be at the discretion of the Investigator. Patients will be followed for second progression on a subsequent treatment, defined according to local practice, and for survival.

Oversight of safety and tolerability of the randomized period of the study will be provided by an Independent Data Monitoring Committee (IDMC). Safety run-in data will not be included in the analysis of the randomized period; however, during the randomized period, the IDMC will review data from the safety run-in separately as part of the ongoing review process.

A futility analysis is planned to occur when approximately 83 Progression free survival (PFS) events have occurred in the randomized period of the study across both treatment arms, which is predicted to occur approximately 15 months after the start of randomization. The futility analysis will be conducted and reviewed by the IDMC to recommend whether the study should continue. The rationale for conducting a futility analysis is to minimize patient exposure in the event that osimertinib with pemetrexed and cisplatin or carboplatin shows insufficient evidence of an efficacy benefit when compared with osimertinib monotherapy.

The overall study design is shown in Figure 1 below.

Figure 1 Study design



1.3 Number of Subjects

Approximately 30 patients to evaluate the safety and tolerability of osimertinib with platinum (carboplatin or cisplatin) and pemetrexed in safety run-in period.

Approximately 556 patients will be randomized, in a 1:1 ratio (osimertinib with chemotherapy: osimertinib) to this study.

The primary analysis of PFS based on Investigator assessment (according to RECIST 1.1) will occur when approximately 278 PFS events (approximately 50% maturity) and at least 16 months of follow-up after LSI has occurred in the 556 randomized patients. This was initially expected to occur approximately 33 months after the first patient is randomized (under an assumed 15-month exponential recruitment); however, the actual DCO for the primary PFS analysis will be determined such that both criteria are met. If the true PFS hazard ratio (HR) for the comparison of osimertinib with chemotherapy vs. osimertinib monotherapy is 0.68, 278 progression events will provide 90% power to demonstrate a statistically significant difference in PFS at a 5% two-sided significance level. This translates to an improvement in median PFS from 19 months to 28 months, assuming exponential distribution and proportional hazards. The minimum critical HR is 0.79, which translates to an approximate median PFS improvement from 19 months to 24 months.

The key secondary endpoint of overall survival (OS) will be tested in a hierarchical procedure, at the time of the PFS analysis and after the primary PFS analysis when the OS data are approximately 60% mature (approximately 334 death events across both arms). Alpha will be strongly controlled across the two OS analyses; i.e., at the time of the primary PFS analysis and at the final OS analysis, with the overall Type 1 error strongly controlled at 5% (2 sided) for the testing of OS under an O'Brien and Fleming spending rule. Under assumed medians of 40 months and 52 months (HR = 0.77) for osimertinib monotherapy and osimertinib with chemotherapy, respectively, 170 observed events (information fraction of 0.51) are expected at the time of the primary PFS analysis with 2-sided alpha of 0.0034, with the remaining alpha assigned to the final OS analysis (0.0490)

2. ANALYSIS SETS

2.1 Definition of Analysis Sets

2.1.1 Safety run-in period

Safety Analysis Set (SAF)

The safety-run in Safety Analysis Set (Safety run-in SAF) contains all patients allocated to the safety run-in period of this study, who receive at least 1 dose of study treatment (osimertinib, cisplatin, carboplatin, or pemetrexed).

Efficacy, disposition, demography, medical history, concomitant medications, and baseline characteristics will be summarized based on the planned treatment cohort.

Safety data will be summarized based on actual treatment cohort according to the platinum agent received:

- patients who received cisplatin will be summarized under osimertinib + cisplatin + pemetrexed cohort (including patients who didn't receive osimertinib and/or pemetrexed);
- patients who received carboplatin will be summarized under osimertinib + carboplatin + pemetrexed cohort (including patients who didn't receive osimertinib and/or pemetrexed);

Note: Patients who did not receive cisplatin or carboplatin will be summarized under the planned cohort (actual treatment will be assigned as planned); patients who switched from cisplatin to carboplatin (or vice versa) will be summarized under the treatment cohort initially received.

Pharmacokinetic (PK) Analysis Set

Pharmacokinetic Analysis Set is defined as patients in the SAF who have at least 1 measurable PK concentration and no missing doses 7 days prior to the PK sample, without any protocol deviation that affects PK, supported by the relevant date and time of this sample. For each time a PK sample is taken, the dosing data for that day should be recorded and the dosing data for the previous day prior to the sample day as well as the sample day should be recorded.

2.1.2 Randomized period

Full Analysis Set (FAS)

The full analysis set (FAS) includes all randomized patients. The FAS will be used for all efficacy analyses and treatment groups will be compared on the basis of randomized study treatment, regardless of the treatment actually received. This is also known as the Intent to Treat (ITT) analysis set.

Safety Analysis Set (SAF)

The safety analysis set (SAF) consists of all randomized patients (i.e. in the FAS) who receive at least 1 dose of study treatment (receive any of osimertinib, cisplatin, carboplatin, or pemetrexed). Safety data will be summarized, according to the treatment actually received; eg, a patient who is randomized to osimertinib plus chemotherapy but who only received osimertinib will be summarized under the osimertinib monotherapy arm.

Pharmacokinetic Analysis Set

Pharmacokinetic Analysis Set is defined as patients in the SAF who have at least 1 measurable PK concentration and no missing doses 7 days prior to the PK sample, without any protocol deviation that affects PK, supported by the relevant date and time of this sample. For each time a PK sample is taken, the dosing data for that day should be recorded and the dosing data for the previous day prior to the sample day as well as the sample day should be recorded.

CNS Full Analysis Set (cFAS)

The CNS FAS (cFAS) is a subset of the FAS population. It includes all patients who undertook a brain scan in the screening/baseline period, had their scan sent for CNS BICR review and were identified by that review as having non-measurable and/or measurable brain disease at baseline (i.e., at least 1 non-measurable and/or 1 measurable brain lesion noted at baseline).

2.2 Protocol Deviations

Protocol deviations are defined as those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study treatments.

A list of important protocol deviations and any action to be taken regarding the exclusion of patients or affected data from specific analyses are defined in Table 1 & Table 3, and a list of deviations that are regarded as important are defined in Table 2 and Table 4, for the safety run-in period and randomized period respectively.

A complete list of anticipated protocol deviations (including important protocol deviations) will be compiled and finalized prior to database lock.

Summaries of the number and percentage of patients with an important protocol deviation by type of deviation will be provided for the safety run-in period and randomized period separately.

By-patient listings of important protocol deviations will be provided.

2.2.1 Safety run-in period

Table 1 Protocol deviations with action to be taken for analysis for safety run-in period

Protocol Deviation	Act to be taken for analysis
Patient did not receive any study medication	Exclude from safety analysis set

Table 2 Important protocol deviations for safety run-in period

Criteria type	Important Deviations Description
Patient was given incorrect study medication	Analyse "As-treated" for the Safety analysis set.
Inclusion	#5: No pathologically confirmed adenocarcinoma of the lung.
	#6: An absence of locally advanced or metastatic NSCLC, not amenable to curative surgery or radiotherapy.
	#7: No presence of a tumour that harbours an EGFR mutation known to be associated with EGFR TKI sensitivity (including exon 19 deletion and L858R).
	#12: No lesions, not previously irradiated, that can be accurately measured at baseline as ≥ 10mm in the longest diameter (except lymph nodes which must have short axis ≥ 15mm) with computerized tomography (CT) or magnetic resonance imaging (MRI) which is suitable for accurate repeated measurements.

Criteria type	Important Deviations Description
Exclusion	#9: Prior treatment with any systemic anti-cancer therapy for advanced NSCLC not amenable to curative surgery or radiation including chemotherapy, biologic therapy, immunotherapy, or any investigational drug. Prior adjuvant and neo-adjuvant therapies (chemotherapy, radiotherapy, immunotherapy, biologic therapy, investigational agents), or definitive radiation/chemoradiation with or without regimens including immotherapy, biologic therapies, investigational agents are permitted as long as treatment was completed at least 12 months prior to the development of recurrent disease. #10: Prior treatment with an EGFR-TKI. #14 Participation in another clinical study with an investigational product during the 4 weeks prior to Day 1. Patients in the follow-up period of an interventional study are permitted. #2: Past medical history of ILD, drug-induced ILD, radiation pneumonitis that requir d steroid treatment, or any evidence of clinically active ILD #4: Mean resting corrected QT interval (QTc) > 470ms, obtained from 3 electrocardiograms (ECGs), using the screening clinic ECG machine derived QTcF value. #4: Any factors that increase the risk of QTc prolongation or risk of arrhythmic vents such as electrolyte abnormalities.

2.2.2 Randomized period

Table 3 Protocol deviations with a tion to be taken for analysis for randomized period

Protocol Deviation	Act to be taken for analysis
Patient did not receive any study medication	Exclude from safety analysis set
Patient was given incorrect study medication	Analyse “As-randomized” for the FAS. Analyse “As-treated” for the Safety analysis set.

Table 4 Important protocol deviations for randomized period

Criteria type	Important Deviations Description
Inclusion	#5: No pathologically confirmed adenocarcinoma of the lung. #6: An absence of locally advanced or metastatic NSCLC, not amenable to curative surgery or radiotherapy.

Criteria type	Important Deviations Description
	#7: No presence of a tumour that harbours an EGFR mutation known to be associated with EGFR TKI sensitivity (including exon 19 deletion and L858R).
	#12: No lesions, not previously irradiated that can be accurately measured at baseline as ≥ 10mm in the longest diameter (except lymph nodes which must have short axis ≥ 15mm) with computerized tomography (CT) or magnetic resonance imaging (MRI) which is suitable for accurate repeated measurements.
Exclusion	#9: Prior treatment with any systemic anti-cancer therapy for advanced NSCLC not amenable to curative surgery or radiation including chemotherapy, biologic therapy, immunotherapy, or any investigational drug. Prior adjuvant and neo-adjuvant therapies (chemotherapy, radiotherapy, immunotherapy, biologic therapy, investigational agents), or definitive radiation/chemoradiation with or without regimens including immotherapy, biologi therapies, investigational agents are permitted as long as treatment was completed at least 12 months prior to the development of recurrent disease.
	#10: Prior treatment with an EGFR-TKI.
	#14: Participation in another clinical study with an investigational product during the 4 weeks prior to Day 1. Patients in the follow-up period of an interventional study are permitted
	#2: Past medical history of ILD, drug-induced ILD, radiation pneumonitis that required steroid treatment, or any evidence of clinically active ILD.
	#4: Mean resting corrected QT interval (QTc) > 470ms, obtained from 3 electrocardiograms (ECGs), using the screening clinic ECG machine-derived QTcF value.
	#4: Any factors that increase the risk of QTc prolongation or risk of arrhythmic events such as electrolyte abnormalities.
	Patient was previously enrolled in the FLAURA2 safety run-in.
IP Administration/ Study treatment	Patient received / used incorrect investigational product.
	Patient restarted study treatment after experiencing Interstitial Lung Disease (ILD).
	Patient did not receive any study medication but was randomized. Patient randomized to chemotherapy arm did not receive pre-dose medications.
Disallowed medications	Other anticancer agents, investigational agents or radiotherapy (for reason other than bone metastases) which are prohibited while the patient is on study treatment.
Procedures/ Tests	Baseline tumour assessments (RECIST1.1) performed more than 28 days before randomization.

Criteria type	Important Deviations Description
	RECIST scans performed outside of the scheduled window on more than 2 occasions.
	Methods/Procedures for tumour assessment are not compliant with CSP or RECIST1.1.
	Missing RECIST assessments for efficacy.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Responses

For both the safety run-in and randomized period, secondary endpoints include the efficacy assessments ORR, DoR, disease control rate (DCR) and depth of response, derived using RECIST 1.1 assessments based on Investigator evaluation. For the randomized period, the primary endpoints are PFS using Investigator assessment as defined by RECIST 1.1, as well as a sensitivity analysis of PFS using BICR assessment as defined by RECIST 1.1

For all patients, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. It will also be used to determine if and when a patient has progressed in accordance with RECIST and their best objective response to study treatment.

Tumor assessments of the chest and abdomen (including the entire liver and both adrenal glands) will be performed using RECIST 1.1 on images from CT (preferred) or MRI with IV contrast. Baseline radiological tumour assessments are to be performed during screening (no more than 28 days before the randomization/first dose and ideally as close as possible to and prior to the date of randomization or first dos for the safety run-in). Thereafter, tumor assessments will be done at week 7 (±1week), Week 13 (±1week), and then every 12 weeks (±1 week), relative to randomization for the randomized period or relative to first dose for the safety run-in until radiological disease progression per RECIST 1.1. Tumor assessments will be done on this schedule even if a patient discontinues treatment prior to progression or receives other anti-cancer treatment.

If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits. This schedule is to be followed in order to minimise any unintentional bias caused by some patients being assessed at a different frequency than other patients.

From the investigator's review of the imaging scans, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD) or progressive disease (PD), using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions (NL) and depending on the status of their disease compared with baseline and previous assessments. If a patient has had a tumour assessment that cannot be evaluated then the patient will be assigned a

visit response of not evaluable (NE), (unless there is evidence of progression in which case the response will be assigned as PD).

For a patient to be considered as a responder does not require confirmation for the randomized period. However, for the safety run-in period, objective response (per RECIST 1.1 using Investigator assessments) is defined as at least 1 visit response of CR or PR that is subsequently confirmed as CR or PR on a subsequent scan (acquired at least 4 weeks later).

Please refer to section 3.1.3 (Table 7) for the definitions of overall response CR, PR, SD and PD.

3.1.1 Target lesions (TLs) – site investigator data

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (LD), (except lymph nodes which must have short axis ≥ 15 mm) with computed tomography (CT) or magnetic resonance imaging (MRI) and which is suitable for accurate repeated measurements. A patient can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomization will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement. In which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as non-target lesions (NTLs) at baseline, including CNS metastases in the Investigator RECIST 1.1 reads. Measurements are not required for these non-target lesions, but their status should be followed at subsequent visits until RECIST v1.1 progression.

Table 5 TL Visit Responses (RECIST 1.1)

Visit Responses	Description
Complete response (CR)	Disappearance of all TLs. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.

Visit Responses	Description
Not evaluable (NE)	Only relevant in certain situations (i.e. if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not applicable (NA)	No TLs are recorded at baseline.

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to one d.p. before assigning a TL response. For example, 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

Missing TL data

For a visit to be evaluable then all TL measurements should be recorded. However, a visit response of PD should still be assigned if any of the following occurred:

- A new lesion is recorded
- A NTL visit response of PD is recorded
- The sum of TLs is sufficiently increased to result in a 20% increase, and an absolute increase of ≥ 5mm, from nadir even assuming the non-recorded TLs have disappeared (Note: the nadir can only be taken from assessments where all the TLs had a LD recorded).

If there is at least one TL measurement missing and a visit response of PD cannot be assigned, the visit response is NE.

If all TL measurements are missing, then the TL visit response is NE. Overall visit response will also be NE, unless there is a progression of non-TLs or new lesions, in which case the response will be PD.

Lymph nodes

For lymph nodes, if the size reduces to < 10 mm then these are considered non-pathological. However, a size will still be given and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are < 10 mm and all other TLs are 0 mm then although the sum may be > 0 mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

Only CR, PD or NE can follow a CR. If a CR has occurred then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0 mm or < 10mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node LD increases by 20% but remains < 10mm.

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0 mm or < 10mm for lymph nodes) then response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.

- Step 3: If not all lesions meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis > 10mm or the reappearance of previously disappeared lesion) or a new lesion appears, then response will be set to PD.

- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR.

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size ('*x*') above which it cannot be accurately measured should be recorded. If using a value of *x* in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team blinded to treatment assignment prior to DBL. It is expected that a visit response of PD will remain in the vast majority of cases.

TL too small to measure

If a TL becomes too small to measure then this will be indicated as such on the case report form and a value of 5 mm will be entered into the database prior to DBL and used in TL calculations. However a smaller value may be used if th radiologist has not indicated 'too small to measure' on the case report form and has entered a smaller value that can be reliably measured. If a TL response of PD results (at a ubsequen visit) then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) should be recorded as NTLs and should not form part of the TL assessment.

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolization), should be handled in the following way. Once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD, this will remain as a valid response category.

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and if ≤ 1/3 of the TLs have missing

measurements then scale up as described in the 'Scaling' section below. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If, after both steps, PD has not been assigned, then, if appropriate (i.e. if ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters calculated in step 2 should be used, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or < 10 mm for lymph nodes) and the lesions that have been subject to intervention have a value of 0 (or < 10 mm for lymph nodes) recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

At subsequent visits, the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with intervention should be treated as missing and scaled up (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of TL measurements are missing (because of inte vention) then the TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e. if using a value of 0 for missing lesions, the sum of diameters has still increased by 20% or more compared to nadir and the sum of TLs has increased by ≥ 5 mm from nad r).

If ≤ 1/3 of the TL measurements are missing (because of intervention) then the results will be scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters) and this will be used in calculations; this is equivalent to comparing the visit sum of diameters of the non-missing lesions to the nadir sum of diameters excluding the lesions with missing measurements.

Example of scaling

Lesion 5 is missing at the follow-up visit; the nadir TL sum including lesions 1-5 was 74 mm.

The sum of lesions 1-4 at the follow-up is 68 mm. The sum of the corresponding lesions at the nadir visit is 62 mm.

Scale up as follows to give an estimated TL sum of 81 mm:

68 x 74 / 62 = 81 mm

CR will not be allowed as a TL response for visits where there is missing data. Only PR, SD or PD (or NE) could be assigned as the TL visit response in these cases. However, for visits with ≤ 1/3 lesion assessments not recorded, the scaled up sum of TLs diameters will be included when defining the nadir value for the assessment of progression.

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions should be summed and reported as the LD for the lesion that split.

Lesions that merge

If two TLs merge, then the LD of the merged lesion should be recorded for one of the TL sizes and the other TL size should be recorded as 0 cm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs, between CT and MRI this will be considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions should be treated as missing.

3.1.2 Non-target lesions (NTLs) and new lesions – site investigator data.

At each visit, the investigator should record an overall assessment of the NTL response. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the investigator's overall assessment of NTLs as follows:

Table 6 NTL Visit Responses

Visit Responses	Description
Complete response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (< 10 mm short axis).
Progressive disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases, the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.
Not evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

To achieve 'unequivocal progression' on the basis of NTLs, there must be an overall level of substantial worsening in non-target disease such that, even in the presence of SD or PR in TLs, the overall tumour burden has increased sufficiently to merit a determination of disease

progression. A modest 'increase' in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

The finding of a new lesion should be unequivocal: i.e. not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour.

New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.

If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present, but should not overtly affect the derivation.

Symptomatic progression is not a descript r for progression of NTLs: it is a reason for stopping study therapy and will not be included in any assessment of NTLs.

Patients with 'symptomatic progression requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumour assessments where possible until objective diseas progression is observed.

3.1.3 Overall visit response – site investigator data

Table 7 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 7 Overall visit responses

Target	Non-Target	New Lesions	Overall Visit Response
CR	CR or NA	No (or NE)	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	Non-PD or NE or NA	No (or NE)	PR
SD	Non-PD or NE or NA	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD

Target	Non-Target	New Lesions	Overall Visit Response
NE	Non-PD or NE or NA	No (or NE)	NE
NA	CR	No (or NE)	CR
NA	Non-CR/Non-PD	No (or NE)	SD
NA	NE	No (or NE)	NE
NA	NA	No (or NE)	NED

3.1.4 Blinded Independent Central Review

There is no blinded independent central review (BICR) of the safety run-in scans. However, all images, including unscheduled visit scans, will be collected on an ongoing basis and sent to an AstraZeneca-appointed Contract Research Organization (CRO) for QC and storage.

For the randomization period, a planned BICR of all radiological imaging data will be carried out using RECIST version 1.1. All radiological scans of randomized patients (including those at unscheduled visits, or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca-appointed CRO to enable BICR up to the point of the primary PFS analysis data cut-off (DCO). After the primary PFS analysis DCO, images will no longer be collected centrally.

Up to three qualified radiologists will independently review all imaging scans in the following way. First a primary review will be performed by two independent radiologists for each patient, on a time point by time point basis to give an overall tumour assessment at each time point using RECIST 1.1. Then a global radiology review will be performed whereby the same independent radiologists will globally assess all time points for a patient in the review period and adjust an overall assessment if necessary. F nally, if the overall assessment for at least one time point for a patient does not agree between the tw independent radiologists, a third independent radiologist will adjudicate and identify which radiologist's assessments they agree with and should be used (for the whole patient). Each scan will be reviewed by 2 independent radiologists using RECIST 1.1 criteria and will be adjudicated if required. The independent reviewers will be blinded to treatment.

For each patient, the BICR will define the overall visit response data (CR, PR, SD, PD, or NE) and the relevant scan dates for each time point (i.e. for visits where response or progression is not identified). If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is any evidence of progression in which case the response will be assigned as PD). Progression free survival will be derived from the overall visit response data and the scan dates.

All CNS scans will have a CNS BICR using modified CNS RECIST guidelines, which allow the selection of up to 5 lesions in the brain as target lesions. The CNS BICR is separate from the BICR and is comprised of independent neuroradiologists.

Further details of the BICR and CNS BICR will be documented in the Independent Review Charter (also referred to as "Imaging Charter").

3.1.5 Imaging of brain metastases

For all patients in either the safety run-in or the randomized period, brain scans should be performed (preferably with MRI) at screening and progression using the same modality. The presence of brain metastases at screening will be based on the Investigator assessment of the brain scan. Patients who have brain metastases or a history of brain metastases at screening are to be followed-up at every imaging visit, and will have tumor assessments according to RECIST 1.1 until overall PD. If brain metastases are not detected (within 28 days prior to randomization [or first dose for the safety run-in]), or if the patient does not have a history of brain metastases, the patient is not required to undergo further imaging of the brain unless metastases are suspected by the Investigator, or until (extracranial) PD is assessed with RECIST 1.1 by the Investigator. Once PD is assessed in patients without brain metastases, a brain scan should be performed within 4 weeks, but preferably as soon as possible, to allow the assessment of new lesions in the brain. If an unscheduled imaging assessment was performed, for example due to suspected CNS progression, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled scan. For patients without a history of brain metastases, the scheduled brain scan will be performed once extracranial progression has been assessed.

3.2 Efficacy Variables

3.2.1 Progression free survival (PFS)

PFS is defined as the time from randomization or from the first dose of study treatment for the safety run-in until the date of objective disease progression or death (by any cause in the absence of progression), regardless of whether the patient withdraws from study treatment or receives another anti-cancer therapy prior to progression. Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST assessment (wi h results CR, PR, SD).

However, if the patient progresses or dies immediately after two or more consecutive missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the two missed visits (Note: NE visit is not considered as missed visit).

If the patient has no evaluable visits or does not have evaluable baseline RECIST 1.1 data they will be censored at study day 1 unless they die within two scheduled RECIST assessments of baseline (12 weeks plus 1 week allowing for a late assessment within the visit window from the randomization date or the first dose date for the safety run-in), in which case their date of death will be used as an event.

Tumor assessments will be done at week 7 (±1week), week 13 (±1week), and then every 12 weeks (±1 week), relative to randomization for the randomized period or relative to the first dose for the safety run-in until radiological disease progression per RECIST 1.1.

If the previous RECIST assessment is at week 7 (study day less than or equal to 43 days plus 1 week window) then two missing visits will equate to 20 weeks since the week 7 assessment, allowing for early and late visits (6 weeks + 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks). If the previous RECIST assessment is at week 13 or after

week 13 (study day >50 and less than or equal to 92 (85 days plus 1 week window) for on week 13, else study day > 92 for after week 13) then two missing visits will equate to 26 weeks since the previous RECIST assessment, allowing for early and late visits (2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks).

The PFS time will always be derived based on scan/assessment dates not visit dates.

RECIST assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- For BICR assessments, the date of progression will be determined based on the earliest of the scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of the reviewer who read baseline first if there is no adjudication for BICR data.

- For investigational assessments, the date of progression will be determined based on the earliest of the dates of the component that triggered the progression.

- For both BICR and investigational assessments, when censoring a patient for PFS the patient will be censored at the latest of the dates contributing to a particular overall visit assessment.

Note: for TLs only the latest scan date is recorded out of all scans performed at that assessment for the TLs and similarly for NTLs only the latest scan date is recorded out of all scans performed at that assessment for the NTLs.

3.2.2 Overall survival (OS)

Overall survival is defined as the time from the date of randomization, or from the first dose of study treatment for the safety run-in, until death due to any cause regardless of whether the patient withdraws from study t eatment or receives another anti-cancer therapy (e.g. date of death or censoring – date of randomization + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Survival calls will be made in the 1 week following the date of the DCO for each OS analysis and for the planned futility analysis. If the death date is after the data cut-off date, these patients will be censored at the date of the data cut-off. Death dates may be found by checking publicly available death registries.

For any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since the SURVIVE module is only completed for patients off treatment if a survival sweep is not performed). The last date for each individual patient is defined as the latest among the following dates (Note: complete dates without imputation) recorded on the case report forms (CRFs):

- AE start and stop dates

- Admission and discharge dates of hospitalization
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anticancer treatment
- Date last known alive on survival status CRF

If a patient is known to have died where only a partial death date is available, then the date of death will be imputed as the latest of the last date known to be alive +1 from the database and the death date using the available information provided:

a. For Missing day only – using the 1st of the month if the month is after the last known alive date, otherwise using the last known alive date + 1
b. For Missing day and Month – using the later of (last known alive date + 1, 1st of January of the year)

If there is evidence of death but the date is entirely missing, it will be treated as missing, i.e. censored at the last known alive date.

3.2.3 Objective response rate (ORR)

Safety run-in period

Confirmed ORR (per RECIST 1.1 using Investigator assessments) is defined as the number (%) of patients with at least 1 visit r sponse of CR or PR, where each CR or PR must be subsequently confirmed at least 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit, based on all treated patients in safety run-in.

Data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue treatment without progression, receive a subsequent anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR calculation (i.e. both visits contributing to a response must be prior to subsequent therapy for the patient to be considered as a responder), where the denominator will be the safety run-in SAF.

In the case where a patient has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the patient is defined as a responder. Similarly, if a patient has visit responses of CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR will be assigned.

If a patient has visit responses of CR, NE, PR, then, as long as the time between the 2 visits of CR/PR is greater than 4 weeks, then a best response of PR will be assigned. Similarly, if a patient has visit responses of PR, NE, CR, then, as long as the time between the 2 visits of PR/CR is greater than 4 weeks, then a best response of PR will be assigned.

Randomized period

ORR (per RECIST 1.1 using Investigator assessments) is defined as the percentage of patients with at least one investigator-assessed visit response of CR or PR and will be based on all randomized patients. ORR will also be defined using the BICR data.

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue randomized treatment without progression, receive a subsequent anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR calculation, where the denominator will be the FAS.

3.2.4 Duration of response (DoR)

Safety run-in period

DoR is defined as the time from the date of first documented response (which is subsequently confirmed) until date of documented progression or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as th latest of the dates contributing towards the first visit that was CR or PR that was subsequently confirmed.

If a patient does not progress following a response, then their DoR will use the PFS censoring time.

Randomized period

DoR is defined as the time from the date of first documented response until date of documented progression or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as the latest of the dates contributing towards the first visit response of PR or CR.

If a patient does not progress following a response, then their DoR will use the PFS censoring time.

3.2.5 Disease control rate (DCR)

Disease control rate (DCR) is defined as the percentage of patients who have a best overall response of CR or PR or SD by RECIST 1.1 as assessed by the Investigator. For patients with a best overall response of SD, a RECIST assessment of SD must have been observed at least 6 weeks minus 1 week to allow for an early assessment within the assessment window (study day 35) following randomization or the first dose for the safety run-in to be included in the

numerator of the calculation for disease control rate. This is to enable sufficient follow-up to establish SD.

3.2.6 Best overall response (BoR)

Safety run-in period

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment. It is the best response a patient has had following the first dose but prior to starting any subsequent cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorization of BoR will be based on RECIST using the following response categories: CR, PR, SD, PD and NE.

CR or PR must be confirmed. For determination of a best response of SD, SD should be recorded at least 6 weeks (can be "at least 5 weeks" as the protocol allows 1 week earlier assessment within the assessment window, i.e. at least 35 days) after first dose.

For patients who die with no evaluable RECIST assessments, if the death occurs ≤13 weeks (i.e. 12 weeks + 1 weeks to allow for a late assessment within the 2 assessment windows) after the first dose but prior to starting any subsequent cancer therapy, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >13 weeks after the first dose then BoR will be assigned to the NE category.

Randomized period

BoR is the best response a patient has had following randomization but prior to starting any subsequent cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorization of BoR will following response categories: CR, PR, SD, PD and NE.

SD should be recorded at least 6 weeks minus 1 week, i.e. at least 35 days (to allow for an early assessment within the assessment window) after randomization.

For patients who die with no evaluable RECIST assessments, if the death occurs ≤13 weeks (i.e. 12 weeks + 1 weeks to allow for a late assessment within the 2 assessment windows) after the randomization but prior to starting any subsequent cancer therapy, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >13 weeks after the randomization then BoR will be assigned to the NE category.

BoR will also be defined using the BICR data.

3.2.7 Depth of Response

Depth of response (i.e., tumour shrinkage / change in tumour size) by Investigator is defined as the relative change in the sum of the longest diameters of RECIST target lesions at the nadir in the absence of new lesions or progression of non-target lesions compared to baseline.

The absolute change and percentage change from baseline in the sum of tumour size at each assessment will be calculated. Tumour size is the sum of the longest diameters of the TLs. The best change in tumour size is the largest decrease from baseline or the smallest increase from baseline in the absence of a reduction. The best change in tumour size will include all assessments prior to progression, or start of subsequent anti-cancer therapy, or the last evaluable RECIST assessment if the patient has not died, progressed or started subsequent anti-cancer therapy.

If the best percentage change cannot be calculated due to missing data (including if the patient has no TLs at baseline), a value of +20% will be imputed as the best percentage change from baseline in the following situations (otherwise best percentage change will be left as missing):

- If a patient has no post-baseline assessment and has died
- If a patient has new lesions or progression of NTLs or TLs
- If a patient has withdrawn due to PD and has no evaluable TL data before or at PD

3.2.8 Post Progression Outcomes

Time from randomization to second progression or death (PFS2)

Time from randomization to second progression or death (PFS2) is defined as the time from the date of randomization to the earliest of the progression event subsequent to first subsequent therapy or death. The second progression event must have occurred after discontinuation of study treatment and following administration of subsequent treatment after the initial PFS event. Patients alive and for whom a second disease progression has not be observed should be censored at the last evaluable RECIST 1.1 assessment.

Time to first subsequent therapy or death (TFST)

Time to first subsequent therapy (TFST) or death is defined as the time from the date of randomization to the earlier of the date of anti-cancer therapy start date following IP discontinuation or death. Any patient not known to have had a subsequent therapy or not known to have died at the time of the analysis will be censored at the last known time to have not received subsequent therapy; i.e., the last follow-up visit where this was confirmed.

Time to second subsequent therapy or death (TSST)

Time to second subsequent therapy (TSST) or death is defined as the time from the date of randomization to the earlier of the date of second subsequent anti-cancer therapy start date following IP discontinuation or death. Any patient not known to have died at the time of the analysis and not known to have had a second subsequent therapy will be censored at the last known time to have not received second subsequent therapy, i.e., the last follow-up visit where this was confirmed. If a patient terminated the study for reason other than death before second subsequent therapy, these patients will be censored at the earliest of their last known to be alive and termination dates.

3.2.9 Central Nervous System (CNS) RECIST 1.1 efficacy variables

Each patient, regardless of presence of CNS metastases at baseline, will undergo a brain scan at RECIST 1.1 progression per the primary PFS endpoint.

Table 8 CNS RECIST 1.1 efficacy endpoints for each group

Groups	Endpoints
Patients in cFAS (with measurable /non-measurable brain disease at Baseline)	By BICR: CNS PFS; CNS ORR; CNS DoR; CNS DCR; Best percentage change in CNS tumour size (target lesion) by BICR
Patients with/without history of CNS metastases (FAS)	CNS PFS by BICR The number of patients with/without CNS metastases at RECIST progression

CNS Endpoints for patients in cFAS

The main exploratory endpoint for the CNS analysis is CNS PFS by CNS BICR.

CNS PFS is defined as the time from randomization until the date of objective CNS progression or death (by any cause in absence of CNS progression) regardless of whether the patient withdraws from randomized therapy or receives another anti-cancer therapy prior to CNS progression. For patients who have not progressed in the CNS or died at the time of analysis ,the same censoring rules will be applied as for the primary PFS.

CNS BICR defined the overall visit response for CNS metastases as complete response (CR), partial response (PR, for TLs only), stable disease (SD, for TLs only), progressive disease (PD), Non-CR/Non-PD (NN, for presence of NTLs only) or not evaluable (NE) at the relevant scan dates for each time point.

CNS ORR is defined as the number (%) of patients who had at least 1 visit with CNS response of PR (TL only) or CR by CNS BICR assessment where the denominator will be cFAS. Data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of CNS ORR. Patients who discontinue treatment without CNS progression, receive a subsequent therapy, and then respond will not be included as responders in the CNS ORR calculation.

CNS DoR is defined as the time from the date of first documented CNS response of PR or CR by CNS BICR assessment until the date of objective CNS progression or death (by any cause in absence of CNS progression) regardless of whether the patient withdraws from randomized therapy or receives another anti-cancer therapy prior to CNS progression. Patients who have not progressed in the CNS or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable brain scan.

CNS DCR is defined as the percentage of patients who have a best CNS BICR overall response of CR or PR (TL only), SD (TL only), or Non-CR/Non-PD (NTL only) by CNS RECIST 1.1 as

assessed by BICR, where the denominator will be cFAS. For patients with a best CNS BICR overall response of SD, a RECIST assessment of SD must have been observed at least 6 weeks (can be "at least 5 weeks" as the protocol allows 1 week earlier assessment within the assessment window) following randomization to be included in the numerator of the calculation for disease control rate.

The best percentage change in CNS TLs tumour size will be defined and calculated in the same way as the depth of response specified in section 3.2.7.

CNS Endpoints for patients with/without history of CNS metastases (FAS)

Subjects in FAS (including subjects not in cFAS) have the same CNS PFS definition as subjects in cFAS.

3.3 Patient Reported Outcome (PRO) Variables

The following PROs will be collected during the randomized period of the study until PFS2 or up to the DCO for the primary PFS analysis, whichever i sooner: EORTC QLQ-C30, EORTC QLQ-LC13, PRO-CTCAE, PGIS, and EQ-5D-5L. All PRO analyses will be based on the full analysis set (FAS). EORTC QLQ-C30 and EORTC QLQ-LC13 are secondary outcomes; all other PROs are exploratory outcomes.

3.3.1 EORTC QLQ-C30 and EORTC QLQ-LC13

Symptoms and overall quality of life will be assessed using EORTC QLQ-C30 and QLQ-LC13. Questionnaires will be scored according to published guidelines or the developer's guidelines, if published guidelines are not available.

The EORTC QLQ-C30 is a valid and reliable PRO instrument in this patient population that was developed by the EORTC Quality of Life Group 1993. It consists of 30 questions that can be combined to produce 3 symptom scales (fatigue, pain, and nausea/vomiting), 5 individual symptom items (dyspnea, insomnia, appetite loss, constipation, and diarrhea), 5 functional scales (physical, role, cognitive, emotional, and social), and a global measure of health status/QoL.

Each subscale, with the number of questionnaire item(s) and item range is shown in Table 9.

Table 9 EORTC QLQ-C30 Scoring version 3.0

	Subscale name	Individual Questionnaire	Number of Items	Item Range
Functioning scales	Physical functioning (PF2)	1 - 5	5	3
	Role functioning (RF2)	6, 7	2	3
	Cognitive functioning (CF)	20, 25	2	3
	Emotional functioning (EF)	21 - 24	4	3
	Social functioning (SF)	26 - 27	2	3
Symptom scales	Fatigue (FA)	10, 12, 18	3	3
	Pain (PA)	9, 19	2	3

	Subscale name	Individual Questionnaire	Number of Items	Item Range
	Nausea/vomiting (NV)	14, 15	2	3
Individual items	Dyspnea (DY)	8	1	3
	Insomnia (SL)	11	1	3
	Appetite loss (AP)	13	1	3
	Constipation (CO)	16	1	3
	Diarrhea (DI)	17	1	3
Global Health Status/QoL		29, 30	2	6
Financial difficulties (FI)		28	1	3

The EORTC QLQ-LC13 is a lung-cancer specific module comprising 13 questions to assess cough, hemoptysis, dyspnea, site specific pain, sore mouth, dysphagia, peripheral neuropathy, and alopecia and pain medication (Bergman et al 1994). With the exception of a multi-item scale for dyspnea, all are single items (Table 10).

Table 10 EORTC QLQ-LC13 Scoring

	Subscale name	Individual Questionnaire	Number of Items	Item Range
Lung cancer symptoms	Coughing (LCCO)	1	1	3
	Hemoptysis (LCHA)	2	1	3
	Dysphagia (LCDS)	7	1	3
	Pain in arm or shoulder	11	1	3
	Pain in other parts (LCPO)	12	1	3
	Pain in chest (LCPC)	10	1	3
Treatment-related symptoms	Sore mouth (LCSM)	6	1	3
	Dyspnea (LCDY)	3,4,5	3	3
	Peripheral neuropathy (LCPN)	8	1	3
	Alopecia (LCHR)	9	1	3

An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/symptom items, the functional scales and the global health status/QoL scale according to the EORTC QLQ-C30 Scoring Manual and QLQ-LC13 instructions, respectively.

The principle for scoring these scales is the same in all cases:

- Step 1: Estimate the average of the items that contribute to the scale; this is the raw score, calculated as $RawScore = RS = (I_4 + I_2 + \ldots + I_n)/n$ where I_n is the score of individual item n.

- Step 2: Use a linear transformation to standardise the raw score, so that scores range from 0 to 100:

  QLQ-C30 Functional scales: $S = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$

  QLQ-C30 Symptom scales / single items / QLQ-LC13 scales: $S = \left\{\frac{(RS-1)}{range}\right\} \times 100$

  QLQ-C30 Global health status / QoL: $S = \left\{\frac{(RS-1)}{range}\right\} \times 100$

  Range is the difference between the maximum possible value of RS and the minimum possible value. The QLQ-C30 and QLQ-LC13 have been designed so that all items in any scale take the same range of values. Therefore, the range of RS equals the range of the item values. Most items are scored 1 to 4, giving range = 3. The exceptions are the items contributing to the global health status / QoL, which are 7 point questions with range = 6.

QLQ-C30 higher scores on the global health status/QoL and functioning scales indicate better health status/function, but higher scores on symptom scales/items represent greater symptom severity; higher scores on QLQ-LC13 represents greater symptom severity.

Changes in score compared with baseline will be evaluated. For each subscale, if < 50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers et al 2001). If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing. The reason for any missing questionnaire will be identified and recorded.

The primary PRO measures will be patient-reported lung cancer symptoms, physical function and global health status/QoL assessed using the EORTC QLQ-C30 and EORTC QLQ-LC13 , namely:

- Dyspnea (3 items scale in EORTC QLQ-LC13)
- Cough: (1 item in EORT QLQ-LC13)
- Chest pain: (1 item in EORT QLQ-LC13)
- Fatigue (3 items scale in EORT QLQ-C30)
- Appetite loss (1 item scale in EORT QLQ-C30)
- Physical function (5 items scale in EORT QLQ-C30)
- Global health status/QoL (2 items scale in EORT QLQ-C30)

Definition of clinically meaningful changes

Changes in score compared to baseline will be evaluated. A minimum clinically relevant change is defined as a change in the score from baseline of ≥ 10 for scales/items from the QLQ-C30 and the QLQ-LC13 (Obosa et al 1998). For example, a clinically relevant deterioration or worsening in chest pain (as assessed by QLQ-LC13) is defined as an increase in the score from baseline (defined as Day 1, pre-dose) of ≥ 10. A clinically relevant improvement in fatigue (as assessed by QLQ-C30) is defined as a decrease in the score from baseline of ≥ 10. At each post-baseline

assessment, change in symptoms/functioning from baseline will be categorized as improved, stable, worsening, or missing as shown in Table 11. Patients with no baseline data or no post-baseline PRO assessment will be excluded from any change from baseline analyses.

Table 11 Mean change and assessment response in symptoms and health-related quality of life

Score	Change from baseline	Visit response
EORTC QLQ-LC13/QLQ-C30 symptom scales/individual items	≥ +10	Worsening
	≤ -10	Improvement
	Otherwise	No change
	Missing	Missing
EORTC QLQ-C30 functional scales global health status/QoL	≥ +10	Improvement
	≤ -10	Worsening
	Otherwise	No change
	Missing	Missing

Time to deterioration (TTD) in symptoms, function and global health status/QoL (QLQ-C30 and QLQ-LC13):

For each of the symptoms, function, or GHS/QoL scales in EORTC QLQ-C30 and each of symptoms in EORTC QLQ-LC13, time to deterioration (TTD) is defined as the time from randomization until the date of the first clinically meaningful worsening (a change in the score from baseline of ≥10) that is confirmed at a subsequent assessment or death (subsequent worsening or death by any cause at least two weeks after the first worsening) in the absence of a clinically meaningful symptom, function, or global health status/QoL worsening, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom, function or GHS/QoL deterioration. Patients with a single worsening and no further assessments will be treated as deteriorated in the analysis. Death will be included as an event only if the death occurs within two assessments of the last PRO assessment where the change could be evaluated. Details of censoring for TTD are presented in below table (Table 12).

Table 12 Censoring rules for TTD

Situation	Event or Censored	Event Date/ Censored Date
No baseline score, or no post-baseline score, or baseline scores of <10 for global health status/QoL and functioning, baseline scales of >90 for symptom scales, and didn't die within two visits of baseline	C	Randomization Date (Study day 1)

Situation	Event or Censored	Event Date/ Censored Date
No baseline score, or no post-baseline score, or baseline scores of <10 for global health status/QoL and functioning, baseline scales of >90 for symptom scales, and die within two visits of baseline	E	Death Date
Confirmed worsening (confirmed by subsequent PRO or followed by death) immediately after two or more consecutive missed visits	C	Latest evaluable PRO assessment prior to the two or more missed visits
Confirmed worsening (confirmed by subsequent PRO or followed by death), without two or more consecutive missed visits	E	The first worsening date
Death (not worsening at last PRO prior to death) immediately after two or more consecutive missed visits	C	Latest evaluable PRO assessment prior to the two or more missed visits
Death (not worsening at last PRO prior to death) without two or more consecutive missed visits	E	Death date
Single worsening at last PRO assessment (no other worsening prior to this last PRO)	E	Last PRO (the single worsening date)
No confirmed worsening or death at the time of analysis	C	Latest evaluable PRO assessment date

Windows for missing two consecutive visits depends on the visit (Analysis visit per Table 21 or Table 22 correspondingly) of last assessment prior to deterioration (prior to death if no deterioration) is listed as below:

- If the last QLQ-LC13 assessment is before or on cycle 3 day, then two scheduled assessments window will be equal to 2 weeks plus 2 days.
- If the last QLQ-LC13 assessment is on cycle 3 day 8, then two scheduled assessments window will be equal to 2 weeks plus 4 days (1 week + 1 day for missed early visit and 1 week + 3 days for missed later visit).

- If the last QLQ-LC13 assessment is on cycle 3 day 15, then two scheduled assessments window will be 4 weeks plus 6 day (1 week + 3 day for missed the early visit and 3 weeks + 3 days for the later visit).
- If the last QLQ-LC13 assessment is on or after cycle 4 day 1, then two scheduled assessments window will be 6 weeks plus 6 days. If a patient dies within two assessments of QLQ-LC13 baseline assessment (2 weeks + 2 days).
- If the last QLQ-C30 assessment is before or on cycle 2 day 1, then two scheduled assessments window will be equal to 6 weeks plus 4 days (3 weeks + 1 day for early visit and 3 weeks + 3 days for the later visit).
- If the last QLQ-C30 assessment is on cycle 3 day 1, then two scheduled assessments window will be equal to 9 weeks plus 6 day (3 weeks + 3 days for early visit and 6 weeks + 3 days for later visit).
- If last QLQ-C30 assessment is on or after cycle 4 day 1 then two scheduled assessments will be 12 weeks plus 6 days. If a patient die within two assessments of QLQ-C30 baseline assessment (6 weeks + 2 days).

Patients whose symptoms, function or global health status/QoL have not shown a clinically meaningful worsening, and alive at the time of the analysis or die after two missed scheduled assessments of last post baseline PRO assessment, will be censored at the time of their last PRO assessment that could be evaluated.

Patients with no baseline score, or with no post-baseline score, or baseline scores of <10 for global health status/QoL and functioning, baseline scales of >90 for symptom scales will be censored at day 1.

Time to definitive deterioration (TTDD) in symptoms, function and global health status/QoL (QLQ-C30 and QLQ-LC13):

Deterioration is defined as definitive if a deterioration is also observed at all subsequent non-missing visits, or a single deterioration followed by death, or a single deterioration followed by monotone missing data afterwards (missed one more more PRO assessments after the single deterioration). Accordingly, time to the definitive deteriation (TTDD) is defined as from the date of randomization to the date at which the subject experiences definitive deterioration. If a subject's best score occurs at more than one assessment, then the earliest potential date will be used as the start date.

Patients with no definitive deterioration will be censored at the time of last PRO assessment. A single worsening with no further assessments will not be considered as definitive deterioration. Death will not be considered as an event for TTDD.

Same as TTD, patients with no baseline score, no post-baseline score, baseline scores of <10 for global health status/QoL and functioning, baseline scales of >90 for symptom scales will be censored at day 1. Details of censoring for TTDD are presented in below table (Table 13)

Table 13 Censoring rules for TTDD

Situation	Event or Censored	Event Date/ Censored Date
No baseline score, no post-baseline score, or baseline scores of <10 for global health status/QoL and functioning, baseline scales of >90 for symptom scales	C	Randomization Date (Study day 1)
Definitive worsening immediately after two or more consecutive missed visits	C	Latest evaluable PRO assessment prior to the two or more missed visits
Definitive worsening without two or more consecutive missed visits	E	The first worsening date
No definitive worsening (including single worsening at last PRO, no two or more consecutive worsening up to last PRO, death without worsening at last PRO prior to death) at the time of analysis	C	Latest evaluable PRO assessment date

In TTD and TTDD analysis, RECIST 1.1 progression will not be considered as deterioration and data will not be affected by RECIST 1.1 progression.

3.3.2 PGIS

The PGIS is a 1- item scale that assesses how a patient perceives his/her overall current severity of cancer symptoms. Patients will choose from response options from "No symptoms" to "Very severe."

3.3.3 EQ-5D-5L

The EQ-5D is a standardized measure of health status developed by the EuroQol Group to provide a simple, generic measure of health for clinical and economic appraisal (EuroQol Group 1990). Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care as well as in population health surveys. The questionnaire assesses 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The patient will be asked to indicate his/her current health state by selecting the most appropriate level in each of the 5 dimensions.

For each dimension, respondents select which statement best describes their health on that day from a possible 5 options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems). A unique EQ-5D health state is referred to by a 5-digit code allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the 5 dimensions.

The EQ-5D profile will be converted into a weighted health state utility value, termed the EQ-5D index, by applying a country-specific equation to the EQ-5D-5L profile that represents the comparative value of health states. This equation is based on national valuation sets elicited from the general population and the base case will be the UK perspective. Where a valuation set has not been published, the EQ-5D-5L profile will be converted to the EQ-5D index using a crosswalk algorithm (Van Hout et al. 2012).

In addition to the descriptive system, respondents will also assess their health on the day of assessment on a visual analogue scale, ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately.

3.3.4 PRO-CTCAE

The PRO-CTCAE will be administered only in countries where a linguistically validated version exists.

The PRO-CTCAE was developed by the National Cancer Institute (NCI) in recognition that collecting treatment-related symptom data directly from patients can improve accuracy and efficiency. This was based on findings from multiple studies demonstrating that physicians and nurses underestimate symptom onset, frequency, and severity in comparison with patient ratings. To date, 78 symptoms of the PRO-CTCAE (version 4) have been identified to be amenable to patient reporting, but not all items are administered in any clinical study. Response options vary in frequency, severity, and interference with usual activities. For this study, 9 symptoms are considered relevant for this cance treatment: mouth or throat sores, nausea, vomiting, loose or watery stools, pain in the abdomen, loss of control of bowel movements, dry skin, hair loss, and numbness or tingling in hands or feet.

3.3.5 Compliance

Compliance and evaluability rate will be calculated for each PRO separately (EORTC QLQ-C30, EORTC QLQ-LC13, PGIS, EQ-5D-5L, PRO-CTCAE):

Compliance rate = number of evaluable forms/number of expected forms × 100

Evaluability rate = number of evaluable forms/number of received forms × 100

- An expected form = a questionnaire that is expected to be completed at a scheduled assessment time, i.e., a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time and not experienced a PFS2 event, excluding patients in countries with no available translation.

- An evaluable form = a questionnaire with a completion date and at least 1 subscale that is non-missing.

- A received form = a questionnaire that has been received and has a completion date and at least 1 individual item completed.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the time point, divided by number of patients still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires, divided by the number of received questionnaires.

3.4 Health Care Resource Use Variables

Health resource utilization will be assessed in terms of hospitalization, outpatient visits and emergency department visits.

3.5 Safety Variables

Safety and tolerability will be assessed in terms of AEs, deaths, laboratory data, vital signs (pulse and blood pressure [BP]), ECG, LVEF, physical exam, and WHO performance status. These will be collected for all patients who take at least one dose of study treatment.

3.5.1 Exposure and dose interruptions/delay

Exposure of Osimertinib

Duration of exposure is defined as follows:

Total (or intended) exposure (month) of study treatment

- Total (intended) exposure (month) = (min(last dose date where dose > 0 mg, date of death, date of DCO) – first dose date +1) / 30.4375

Actual exposure time (months) will be calculated from first dose to the last dose, taking account of dose interruptions

- Actual exposure = (intended exposure – total duration of dose interruptions)/30.4375

  Where intended exposure will be calculated as above and a dose interruption is defined as any length of time where the patient has not taken any of the planned daily dose.

The total and actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Exposure of chemotherapy (Pemetrexed, Cisplatin, and Carboplatin)

Duration of treatment on chemotherapy will be in terms of the number of cycles and total exposure.

A cycle corresponds to a period of 21 days. If a cycle is prolonged due to toxicity, this should still be counted as one cycle. A cycle will be counted if chemotherapy is started, even if the full dose is not delivered. The total exposure (month) will be calculated as the (min(last dose date where dose > 0 mg for any of Pemetrexed, Cisplatin, and Carboplatin, date of death, date of DCO) – first dose date + 21) / 30.4375.

Missed or forgotten doses

Missed and forgotten doses should be recorded on the DOSE module as a dose interruption with the reason recorded as "Patient forgot to take dose". Dose interruptions for Osimertinib and chemotherapy and dose delays for chemotherapy will be summarized by each study drug.

Patients who permanently discontinue during a dose interruption

If a patient permanently discontinues study treatment during a dose interruption, then the date of last administration of study medication recorded on DOSDISC will be used in the programming.

Safety Follow-up

- Total Safety Follow-up = min((last dose date +28), date of withdrawal of consent, date of death, date of DCO, date of first subsequent anti-cancer therapy) – first dose date +1

3.5.2 Dose intensity and percentage intended dose for Osimertinib and Pemetrexed

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. Relative dose intensity (RDI) will be defined as follows:

- RDI = 100% * d/D, where d i the actual cumulative dose delivered up to the actual last day of dosing and D is the ntended cumulative dose up to the or the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule.

Actual cumulative dose will be calculated by summing ((End date of study drug administration - Start date of study drug administration + 1) x Dose), for each period of study drug administration recorded on the study drug exposure form up to min(date of last dose date where dose > 0, date of death, date of DCO).

Intended cumulative dose will be calculated by summing the individual doses that should have been received up to and including the last day of day of treatment according to the protocol planned dose and schedule. This is given by: Integer part (rounded up) of ((min(date of last dose date where dose > 0, date of death, date of DCO) – first dose date +1)×intended dose according to the protocol).

The percentage intended dose (PID) will be calculated in the same way as RDI. But D is the intended cumulative dose up to the progression instead of the date of last dose where dose >0. D is therefore given by: Integer part (rounded up) of ((min(date of progression, date of study discontinuation, date of death, date of DCO) – first dose date +1)×intended dose according to the protocol).

3.5.3 Adverse events (AEs)

AEs will be collected from the date of informed consent form throughout the treatment period and including the 28-day follow-up period (28 days after last dose of IP).

Serious AEs will be recorded from the time of signing of informed consent form. SAEs considered related to study treatment and/or study procedures will be collected throughout progression follow-up. SAEs considered related to study treatment will be collected throughout survival follow-up.

Adverse Events is defined as treatment emergent if they onset, or worsen (by investigator report of a change in intensity), from the date of first dose up to and including 28 days after last dose of IP but prior to start of a subsequent anti-cancer treatment. A treatment related AE will be considered as treatment emergent regardless the AE onset date. The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) version 5.0.

A SAE is an AE occurring during any study phase (including treatment and follow-up), that fulfils one or more of the following criteria:

- Results in death
- Is immediately life-threatening
- Requires in-patient hospitalization or prolongation of existing hospitalization
- Results in persistent or significant disability or incapacity
- Is a congenital abnormality or birth defect
- Is an important medical event that may jeopardise the patient or may require medical treatment to prevent one of the outcomes listed above

Other significant adverse events (OAE)

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and 'Discontinuation of Investigational Product due to Adverse Events' (DAEs). Based on the expert's judgement, significant adverse events of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered other significant adverse events (OAEs) and reported as such in the Clinical Study Report (CSR). A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

Examples of these maybe marked hematological and other laboratory abnormalities, and certain events that lead to intervention (other than those already classified as serious) or significant additional treatment.

AEs of special interest

AEs of special interest (AESIs) represent pre-specified risks that are considered to be of importance to a clinical development program.

These AESIs have been identified as a list of categories provided by the patient safety team. Preferred terms used to identify adverse events of special interest will be listed before database lock and documented in the Study Master File. Groupings of certain MedDRA preferred terms will be based on preferred terms provided by the medical team and a listing of the preferred terms in each grouping will be produced. The grouped terms expected are ILD and pneumonitis, Hematological Toxicities, and cardiac effects (cardiac failure).

Other categories may be added as necessary or existing terms may be merged. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AEs of interest and identified which higher-level terms and which preferred terms contribute to each AESI. Further reviews may take place prior to database lock (DBL) to ensure any further terms not already included are captured within the categories.

3.5.4 Analysis of Total Calcium per NCI CTCAE criteria

Corrected calcium records will be programmatically derived from Total Calcium and Albumin and appended to the lab dataset for grading.

Corrected Calcium = serum calcium + 0.8 × (4 - serum albumin)

3.5.5 Laboratory Safety Variables

The following laboratory variables (Table 14) will be summarized:

Table 14 Laboratory safety variables

Hematology/Hemostasis (whole blood [B])	**Clinical Chemistry (Serum [S] / Plasma [P])**	**Urinalysis (Urine [U])**
B-Hemoglobin (Hb)	S/P-Albumin	U-Glucose
B-Red Blood Cell (RBC) count	S/P-Alanine transaminase (ALT)	U-Protein
B-Hematocrit	S/P-Aspartate transaminase (AST)	U-Blood
B-Reticulocytes	S/P-Alkaline phosphatise (ALP)	
B-Leukocyte count	S/P-Bilirubin, total	
B-Leukocyte differential count (absolute count) [a]	S/P-Calcium, total	
Neutrophils	S/P-Corrected Calcium	
Lymphocytes	S/P-Creatinine	
Monocytes	S/P-Creatinine Clearance [b]	
Basophils	S/P-Glucose	
Eosinophils	S/P-Lactate dehydrogenase (LDH) [c]	

Hematology/Hemostasis (whole blood [B])	Clinical Chemistry (Serum [S] / Plasma [P])	Urinalysis (Urine [U])
B-Platelet count	S/P-Magnesium	
	S/P-Potassium	
	S/P-Sodium	
	S/P-Urea/Blood Urea Nitrogen	

a. The value is to be provided as percentage of the leukocyte count if the absolute leukocyte differential counts are not available, and vice versa.
b. Creatinine clearance will be derived using the method of Cockcroft and Gault (Cockcroft & Gault, 1976).
c. LDH is an additional variable collected during screening

3.5.6 Vital Signs

Vital signs will include systolic and diastolic blood pressure weight, and pulse rate.

3.5.7 Electrocardiograms (ECG)

The average of triple ECG results at each timepoint will be derived for the analysis.

Fridericia QTc correction (QTcF) will be calculated as:

$$QTcF = \frac{QT}{\sqrt[3]{RR}}$$

3.5.8 Who performance status

Performance status will be assessed at the scheduled visits indicated in CSP schedule of assessments (SoA) table according to WHO criteria as follows:

0 = Fully active, able to car y out all pre-disease activities without restrictions.

1 = Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light housework, office work.

2 = Ambulatory and capable of self-care but unable to carry out any work activities. Up and about more than 50% of waking hours.

3 = Capable of only limited self-care, confined to bed or chair more than 50% of waking hours.

4 = Completely disabled, cannot carryon self-care, totally confined to bed or chair.

3.5.9 Left ventricular ejection fraction (LVEF)

An echocardiogram or Multi Gated Acquisition Scan (MUGA) scan to assess LVEF will be performed at the visits indicated in CSP SoA.

The modality of the cardiac function assessments must be consistent within a patient; i.e., if echocardiogram is used for the screening assessment, then echocardiogram should also be used for subsequent scans. The patients should also be examined using the same machine and operator whenever possible, and quantitative measurements should be taken. If an on-treatment assessment is abnormal at the time of discontinuation of study therapy, a 28-day follow-up assessment will be required to confirm reversibility of the abnormality. If a patient has had a MUGA scan or echocardiogram performed within 28 days prior to treatment discontinuation, the discontinuation visit Echo/MUGA scan is not required unless clinically indicated.

3.5.10 Physical examinations

Physical examination includes assessments of the following: general appearance, skin, head and neck (including ears, eyes, nose and throat), respiratory, cardiovascular, abdomen, lymph nodes, thyroid, musculoskeletal (including spine and extremities) and neurological systems. Height is measured at screening only.

3.6 Pharmacokinetic Variables

Plasma samples for pharmacokinetic assessments will be collected as per the protocol for both the safety run-in and the randomized portion of the study. Samples will be analysed by Covance laboratories, on behalf of AstraZeneca.

Pharmacokinetic analysis of the plasma concentration data for osimertinib and metabolite AZ5104 will be performed by Covance on behalf of Quantative Clinical Pharmacology. The actual sampling times will be used in PK parameter calculations and PK parameters will be derived using concentration time profile and standard non-compartmental methods. The analysis for safety run-in will be conducted separately from the randomized portion of the study.

Where possible, the following PK parameters will be determined for osimertinib and metabolite AZ5104.

- Maximum plasma concentration at steady state ($C_{max, ss}$)
- Time to $C_{max, ss}$ ($t_{max, ss}$)
- Minimum plasma concentration at steady state ($C_{min, ss}$)
- Area under the plasma concentration-time curve over a dosing interval at steady state (AUC_{ss})
- Apparent plasma clearance at steady state (CL_{ss}/F, osimertinib only)
- Metabolite to parent ratio for AUC_{ss} and $C_{max, ss}$

$C_{max, ss}$, $C_{min, ss}$, and $t_{max, ss}$ will be obtained directly from the plasma concentration-time profile. To calculated AUC_{ss}, the predose sample will also be used as the end of dosing interval sample. CL_{ss}/F will be calculated as Dose/ AUC_{ss}.

For visit based summaries of PK data, summary statistics will be presented where at least 3 observations above lower level of quantification (LLoQ) in a treatment group have data recorded

at a particular visit. Two values are presented as a minimum and maximum with the other summary statistics as non-calculable (NC).

PK parameters for osimertinib and AZ5104 may be estimated from the plasma concentration data from the sparse PK sampling regimen employed by combining with data from other studies by use of population PK modelling techniques. The population PK analysis, if conducted may be reported separately from the CSR.

4. ANALYSIS METHODS

Table 15 summarises the planned summaries and analyses for safety run-in period.

Table 15 Planned summaries and analyse for safety run-in period

Summaries		Analysis Population
Demographics	Demography and baseline	SAF
	Medical history	SAF
	Prior and concomitant treatments	SAF
Efficacy	ORR*, DoR*, Depth of response* DCR*	SAF
	OS	SAF
PK	PK	Pharmacokinetic (PK) Analysis Set
Safety	Adverse Event	SAF
	Laboratory evaluations	SAF
	Vital signs	SAF
	Physical ex mination	SAF
	ECG	SAF
	LVEF	SAF
	WHO performance status	SAF

* Based on investigator assessment using RECIST 1.1.

Table 16 summarises the planned summaries and analyses for randomized period.

Table 16 Planned summaries and analyse for randomized period

Summaries		Analysis Population
Demographics and patient characteristics	Demography and baseline	FAS
	Medical history	FAS
	Prior and concomitant treatments	FAS
	PFS**	FAS

Efficacy (including PROs)	OS	FAS
	ORR**, DoR**, DCR*, Depth of response*	FAS
	EORTC QLQ-C30, QLQ-LC13	FAS
	PGIS, EQ-5D, PRO-CTCAE,	FAS
	PFS2, TFST, TSST	FAS
PK	PK	Pharmacokinetic (PK) Analysis Set
Safety	Adverse Event	SAF
	Laboratory evaluations	SAF
	Vital signs	SAF
	Physical examination	SAF
	ECG	SAF
	LVEF	SAF
	WHO performance status	SAF

* Based on investigator assessment using RECIST 1.1.

** Based on investigator assessment using RECIST 1.1, and r peated using RECIST 1.1 by BICR assessment.

4.1 General Principles

4.1.1 Populations for analyses

Safety run-in period

Safety and efficacy summaries (excluding PK data summaries) in the safety run-in period will be based on the Safety Analysis Set.

Randomized period

Efficacy and Health related quality of life (HRQoL) data will be summarised and analysed based upon the FAS. Safety and treatment exposure data will be summarised based upon the safety analysis set. Study population and demography data will be summarised based upon the FAS. CNS RECIST efficacy endpoints will be analysed based on cFAS.

4.1.2 General principles for safety run-in and randomized period

The below mentioned general principles will be followed for both safety run-in period and randomized period:

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarized by the number of observations, mean, standard deviation, median, upper and lower quartiles minimum, and maximum. For log-transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables will be summarized by frequency counts and percentages for each category.

- Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment group. Overall totals will be calculated for baseline summaries only.

- For continuous data, the mean and median will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.

- For categorical data, percentages will be rounded to 1 decimal place.

- SAS® version 9.3 (or higher) will be used for all analyses.

A month is operationally defined to be 30.4375 days. One year is defined to be 365.25 days.

Data will be presented in data listings by patient identifier and treatment arm or treatment cohort for the safety run-in.

Where analysis models are stratified by the randomization stratification factors, the strata obtained at randomization will be used, not the values recorded in the electronic case report form (eCRF).

4.1.3 Baseline definition and post baseline summaries

In general, for efficacy the last observed measurement prior to randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before the first dose of randomized treatment then this assessment will be used as baseline. PRO assessments on cycle 1 day 1 will be considered as baseline for PRO analysis and summaries.

For safety endpoints the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified. The average of triple ECG results at the last assessment before the first dose will be derive as the baseline. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The percentage change from baseline will be calculated as (post-baseline value - baseline value) / baseline value × 100.

4.1.4 Handling missing data

In general, other than for partial dates, missing data will not be imputed and will be treated as missing with the exceptions specified for certain efficacy variable.

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e. below the lower limit of quantification) or "> x" (i.e. above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings.

Additionally, adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

4.1.4.1 Imputation of partial dates

Should the whole date be missing it is more difficult to follow a general principle and these should be reviewed within the study and decided how to be handled.

Generally, the imputation of dates is used to decide if an observation is treatment emergent for adverse events or concomitant medications. The imputed dates are not advised to be used to calculate durations where the results would be less accurate

- For missing initial diagnostic dates, if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.
- For missing Concomitant medication, radiotherapy and AE start dates, the following will be applied:
    a. Missing day - impute the 1st of the month unless month is the same as month of the first dose of study drug then impu e first dose date
    b. Missing day and month - impute 1st January unless year is the same as first dose date then impute first dose dat
    c. Completely missing - impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date
- For missing concomitant medication, and AE end dates, the following will be applied:
    a. Missing day - impute the last day of the month
    b. Missing day and month - impute 31st December Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.
- If a patient is known to have died where only a partial death date, the imputation is specified in OS section (Section 3.2.2)
- For missing the first subsequent anti-cancer therapy date, the followings will be applied:
    a. Missing day - impute the 1st of the month unless month is the same as month of the last dose of study drug then impute the last dose date + 1 day

b. Missing day and month - impute 1st January unless year is the same as the last dose date then impute the last dose date + 1 day
c. Completely missing - impute the last dose date + 1 day

- For missing the second subsequent anti-cancer therapy date, the followings will be applied if the first subsequent anti-cancer therapy end date is available:

  a. Missing day - impute the 1st of the month unless month is the same as month of the first subsequent anti-cancer therapy end date then impute the first subsequent therapy end date + 1 day
  b. Missing day and month - impute 1st January unless year is the same as the first subsequent anti-cancer therapy end date then impute the first subsequent therapy end date + 1 day
  c. Completely missing – impute the first subsequent therapy end date + 1 day

- For missing first subsequent anti-cancer therapy end dat in the scenario that a patient received the second subsequent therapy with missing start date, then the first subsequent therapy end date will be imputed as following rules, a d the imputed first anti-cancer therapy end date will be used for the missing second subsequent therapy start date imputation:

  a. Missing day - impute the 1st of the mon h unless month is the same as month of the first subsequent anti-cancer therapy start date then impute the first subsequent therapy start date + 1 day
  b. Missing day and month - impute 1st January unless year is the same as the first subsequent anti-cancer th rapy start date then impute the first subsequent therapy start date + 1 day
  c. Completely mis ing - impute the first subsequent therapy start date + 1 day

4.1.4.2 Imputation rules for lab values outside of quantification range

Lab values below the lower limit of quantification (LLoQ) that are reported as "< LLoQ" or "≤ LLoQ" in the database will be imputed by LLoQ × 0.99 for analysis purposes. The original value will be listed.

Lab values above the upper level of quantification (ULoQ) that are reported as "> ULoQ" or "≥ ULoQ" in the database will be imputed by ULoQ × 1.01 for analysis purposes. The original value will be listed.

4.2 Analysis Methods

No formal statistical testing is planned within the safety run-in. At the Safety Review Committee assessment for the safety run-in, a small number of efficacy summaries and analyses will be produced to report on the efficacy data collected to that point.

For randomized period, the primary analysis of PFS based on Investigator assessment (according to RECIST 1.1) will occur when approximately 278 PFS events (approximately 50% maturity)

and at least 16 months of follow-up after LSI have occurred in the 556 randomized patients. If the true PFS HR for the comparison of osimertinib with chemotherapy vs. osimertinib monotherapy is 0.68, 278 progression events will provide 90% power to demonstrate a statistically significant difference in PFS at a 5% two-sided significance level. This translates to an improvement in median PFS from 19 months to 28 months, assuming exponential distribution and proportional hazards. The minimum critical HR is 0.79, which translates to an approximate median PFS improvement from 19 months to 24 months.

Table 17 is a summary of statistical method to be conducted for the randomized period.

Table 17 Statistical analyses to be conducted for randomized period

Endpoints Analysed	Notes
PFS	Primary analysis stratified log-rank test Sensitivity analyses[a] 1. Analysis using alternative censoring rules 2. Covariate adjusted Cox regression Subgroup Analysis
Overall survival	Stratified log r nk test
Objective response rate	Logistic regression model
Duration of response	Kaplan-M ier estimates (no formal comparison or p-value)
Disease control rate	Logistic regression model
PRO	MMRM (Change from baseline) Stratified log-rank test and Kaplan-Meier estimates (Time to deterioration)

[a] *See section 4.2.3.2 for furth r details*

4.2.1 Multiplicity - randomized period

In order to provide strong control of the type I error rate, α=0.05 (two-sided), the primary endpoint PFS and the secondary endpoint OS will be tested in sequential order. If the previous analysis in the sequence is not statistically significant, the alpha will not be transferred to subsequent analyses.

At the time of the primary analysis of PFS, if the PFS analysis is statistically significant, then subsequent hypothesis testings for OS will be performed at overall α=0.05 significance level (two-sided) using O'Brien and Fleming spending function. If the PFS analysis is not statistically significant at the time of the PFS analysis then the hypothesis testings of OS will not be performed.

The key secondary endpoint of OS will be tested in a hierarchical procedure, at the time of the PFS analysis and at the final analysis when the OS data are approximately 60% mature (approximately 334 death events across both arms).

The significance level for the OS analyses will be calculated using the statistical software package EAST® by specifying the information fraction for each analysis. The information fraction is calculated as the number of OS events at the analysis time-point divided by the total number of planned events of 334 at the final analysis time-point. For example, under assumed medians of 40 months and 52 months (HR = 0.77) for osimertinib monotherapy and osimertinib with chemotherapy, respectively, 170 observed events (information fraction of 0.51) are expected at the time of the primary PFS analysis with 2-sided alpha of 0.0034, with the remaining alpha assigned to the final OS analysis (0.0490).

4.2.2 Time-to-event endpoint considerations - randomized period

Log-rank vs Cox

Time-to-event data (e.g. PFS, OS) will be analysed using a stratified log-rank test.

The log-rank test will be stratified according to the values recorded in the randomization system, as the analysis is then consistent with the possible permutations of the randomization.

A Cox proportional hazards model containing treatment and the stratification factor(s) alone will be used to estimate the PFS HR to ensure that output from the Cox model is likely to be consistent with the results of the primary analysis using the stratified log-rank test.

Handling of ties

Efron method will be used to handle ties in Cox proportional model.

Hazard ratio and confidence interval estimation

The primary analysis, progression f e survival per the Investigator assessment for patients in the FAS, will be analysed using a log rank test stratified by race (Chinese/Asian vs. Non-Chinese/Asian vs. Non-Asian) WHO performance status (0 vs. 1), and method used for tissue testing (central vs. local) for generation of the p-value, using the Efron method for handling ties.

If the resulting strata are too small (i.e., < 20 events) the strata will be collapsed in the following pre-defined order to allow analysis. The China cohort strata will be collapsed first (to Asian vs. Non-Asian), followed by race, and then WHO PS, and finally central/local tissue testing.

The HR and CI will be obtained directly from the U and V statistics as follows (Berry et al 1991; Robins et al 1991; Robins 1993; Selke & Siegmund 1983):

$$HR = exp\left(\frac{U}{V}\right)$$

$$95\%\ CI\ for\ HR = \left(exp\left\{\frac{U}{V} - \frac{1.96}{\sqrt{V}}\right\}, exp\left\{\frac{U}{V} + \frac{1.96}{\sqrt{V}}\right\}\right)$$

Where $U = \sum_k U_k = \sum_k \sum_i (d_{1ki} - e_{1ki})$ is the stratified log-rank test statistic (with d_{1ki} and e_{1ki}, the observed and expected events in group 1, stratum k) and $\sqrt{V} = \sqrt{\sum_k V_k}$ is the standard deviation of the log-rank test statistic obtained from the SAS® LIFETEST procedure with a term for the stratification.

PFS by BICR and OS data will be analysed using the same methodology and model as for the primary analysis of PFS by investigator.

Proportionality assumption

The assumption of proportionality will be assessed for PFS by investigator, PFS by BICR, CNS PFS by BICR, and OS. In the event of non-proportionality, the HR will be interpreted as an average HR over the observed extent of follow-up.

Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time dependent covariate (adding a treatment-by-time or treatment-by-ln(time) interaction term) to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HR calculated over distinct time-periods for example 0-6m, 6-12m etc. In such circumstances, the HR from the primary analysis can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found this may be a result of a treatment-by-covariate interaction, which will be investigated.

4.2.3 Progression free survival (PFS)

4.2.3.1 Safety run-in period

PFS by Investigator will not be summarized at the Safety Review Committee safety-run in assessment. A listing of PFS times along with the RECIST responses will be produced. The listing will be updated at the time of the primary PFS DCO. The median PFS with its 2-sided 95% CI will be provided if have ≥ 20 events at the time of primary PFS analysis.

4.2.3.2 Randomized period

The treatment status at progression of patients at the time of analysis will be summarized. This will include the number (%) of patients who were on treatment at the time of progression, the number (%) of patients who discontinued study treatment prior to progression, the number (%) of patients who have not progressed and were on treatment or discontinued treatment. This will also provide distribution of number of days prior to progression for the patients who have discontinued treatment.

The progression status at the time of the PFS analysis will also be summarized, including the number and percentage of patients who progressed due to RECIST progression or due to death, or did not progress / death by categories of PFS censoring reason.

Kaplan-Meier (K-M) curves, K-M estimates 25th, 50th (median), and 75th percentiles along with their 2-sided 95% confidence intervals (CIs) for PFS will be presented by randomized treatment

group. The percentage PFS with 95% CI at specific timepoints (e.g. 6, 12, 18, 24, 36 month etc.) will be summarized.

Details of censoring for primary PFS analysis are presented in below table (Table 18):

Table 18 Censoring rules for primary PFS

Situation	Event or Censored	Event Date/ Censored Date
No evaluable post-baseline visits or does not have baseline RECIST 1.1 data, and didn't die within two visits of baseline	C	Randomization Date (Study day 1)
No evaluable post-baseline visits or does not have baseline RECIST 1.1 data, die within two visits of baseline	E	Death Date
Progresses or dies immediately after two or more consecutive missed visits	C	Latest evaluable RECIST 1.1 assessment prior to the two missed visits
Disease progression or death (by any cause in the absence of progression) without two or more consecutive missed visits regardless of whether the patient withdraws from randomized therapy or receives another anti-cancer therapy prior to progression	E	Disease progression date, or death date if no PD
Not progressed or died at the time of analysis	C	Latest date of assessment from their last evaluable RECIST assessment

4.2.3.3 PFS sensitivity analyse

Randomization bias

A Cox proportional hazards model will be employed to assess the effect of the pre-specified covariates listed in section 4.2.9 on the PFS HR estimate. A model will be constructed, containing treatment and the stratification factors alone, to ensure that any output from the Cox model is likely to be consistent with the results of the primary analysis using the stratified log-rank test. The results from the initial model and the model containing additional covariates (in the MODEL statement) will be presented.

This analysis evaluates the treatment effect, adjusting for any potential imbalances in baseline prognostic factors that are not balanced by stratification.

The model will include all additional covariates listed in section 4.2.9 regardless of whether their inclusion significantly improves the fit of the model, providing there is enough data to make them meaningful. Missing covariate data will be imputed using the mean (for continuous variables) or the most common category (for categorical factors).

Quantitative interaction

The presence of quantitative interactions will be assessed by means of an overall global interaction test. This will be performed in the overall population by comparing the fit (likelihood ratio test for the change in degrees-of-freedom) of a Cox proportional hazards model including treatment, covariates listed in PFS subgroup analysis (Section 4.2.9), and all covariate by treatment interaction terms, with one that excludes the interaction terms and will be assessed at the two-sided 10% significance level. If the fit of the model is not significantly improved (i.e. not statistically significant), then it will be concluded that overall, the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and any newly significant interactions re-included until a final model is reached where all included interactions are significant and all excluded interactions are non-significant. Throughout this process, all main effects will be included in the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

Any quantitative interactions identified using this procedure will then be tested to rule out any qualitative interaction using the approach of Gail and Simon (Gail, M. & Simon, R., 1985).

Ascertainment bias

Ascertainment bias will be assessed by analysing the BICR data. The stratified log rank test will be repeated on PFS using the BICR da a based upon RECIST. The HR and CI will be presented.

If there is an important discrepancy between the primary analysis using the Investigator data and this sensitivity analysis using BICR data, then the proportion of patients with Investigator assessment but no central confirmation of progression will be summarized; such patients have the potential to induce bias in the central review due to informative censoring. An approach of imputing an event at the next visit in the central review analysis may help inform the most likely HR value (Fleischer et al 2011), but only if an important discrepancy exists.

Disagreements between Investigator and central reviews of RECIST overall response (CR, PR, SD, PD) will be presented for each visit and each treatment group.

Evaluation-time bias

A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled time points. The midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be analysed using a stratified log-rank test, as described for the primary analysis of PFS. For patients whose death was treated as a PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this

analysis, the mean of patient-level average inter-assessment times will be tabulated for each treatment. This approach will use the Investigator RECIST assessments.

Table 19 Change of censoring rule for PFS evaluation-time bias sensitivity

Situation	Event or Censored	Event Date/ Censored Date
Progresses or dies immediately after two or more consecutive missed visits	C	The midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment)

Attrition bias

Attrition bias will be assessed by repeating the PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following 2 or more non-evaluable tumour assessments will be included. In addition, and within the same sensitivity analysis, patients who take subsequent anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last evaluable RECIST assessment or progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring using the PFS data from the primary analysis and where the censoring indicator of the PFS analysis is reversed.

Table 20 Change of censoring rules f r PFS attrition bias sensitivity analysis

Situation	Event or Censored	Event Date/ Censored Date
Progresses or dies immediately after two or more consecutive missed visits	E	Disease progression date, or death date if no PD
Take subsequent therapy prior to their last evaluable RECIST assessment or progression or death	C	Last evaluable assessment prior to taking the subsequent therapy

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression free survival.

4.2.4 Overall survival (OS)

4.2.4.1 Safety run-in period

At the time of the primary PFS DCO, Kaplan-Meier (K-M) curves, K-M estimates 25th, 50th (median), and 75th percentiles for OS will be provided for all safety run-in subjects. The percentage OS with 95% CI at specific timepoints (e.g. 6, 12, 18, 24, and 36 month, etc.) will be summarized. A listing of OS data will be provided.

4.2.4.2 Randomized period

The analysis of OS will be conducted at 2 time points: at the time of the primary analysis of PFS and at approximately 60% maturity, when approximately 334 death events (across both arms) have occurred.

The alpha will be split between the two analyses to provide strong control of the family-wise error rate (see Section 4.2.1).

Overall survival data will be analyzed using the same methodology and model as for the analysis of PFS, provided there are sufficient events (≥ 20 deaths) available for a meaningful analysis; otherwise, descriptive summaries will be provided. The percentage OS with 95% CI at specific timepoints (e.g. 6, 12, 18, 24, and 36 month, etc.) will be summarized for the analysis of OS at the primary PFS analysis, and the percentage OS at additional months (e.g., 36 months, 42 months, etc.) will also be summarized for the analysis of OS at the final survival follow up.

An exploratory analysis will be conducted to report the 5 year overall survival rate. The data cut-off will occur 5 years after the last patient has been rand mized or all patients have died, whichever occurs first. No p-value will be presented as this analysis is excluded from the alpha spending rule.

4.2.5 Objective response rate (ORR)

The ORR will be based on the site investigator RECIST data, and using all scans regardless of whether they were scheduled or not.

4.2.5.1 Safety run-in period

The confirmed ORR, along with the 95% exact CI, will be summarized for all safety run-in patients in SAF.

4.2.5.2 Randomized period

ORR by Investigator will be compared between osimertinib plus chemotherapy treatment versus osimertinib treatment using logistic regression models stratified by race (Chinese/Asian vs. Non-Chinese/Asian vs. Non-Asian), WHO PS (0 vs. 1), and method used for tissue testing (central vs. local).

The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour osimertinib plus chemotherapy treatment) together with its associated 95% profile likelihood CI (e.g. using the option 'LRCI' in SAS® procedure GENMOD) and two-sided p-value. The p-value will be based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model that contains the covariates defined above.

If there are not enough responses for a meaningful analysis (at least 20 total responses, with a minimum of 5 response per treatment arm) using logistic regression then a Cochran–Mantel–Haenszel (CMH) test will be used.

Summaries will be produced that present the number and percentage of patients with a tumour response (CR/PR) by stratification factor for randomized period.

Same analysis and summaries will also be produced for ORR by BICR. For each randomized treatment arm, best objective response (BoR) will be summarized by n (%) for each category (CR, PR, SD, PD and NE). No formal statistical analyses are planned for BoR.

The analysis of ORR will take place at the time of the primary PFS analysis only.

4.2.6 Duration of response (DoR)

For safety-run in period, descriptive data will be provided for the DoR by Investigator in responding patients, including the associated swimmers plot (without any formal comparison or p-value attached).

For the randomized period, the same summaries will be provided by randomized treatment group, and will be repeated for DoR by BICR.

Kaplan-Meier (K-M) curves, K-M estimates 25th, 50th (m dian), and 75th percentiles for DoR will be provided (no formal comparison or p-value) for the randomized period.

The analysis of DoR will take place at the time of the primary PFS analysis only.

4.2.7 Disease control rate (DCR)

4.2.7.1 Safety run-in period

DCR by investigator will be summarize along with the 95% exact CI for all safety run-in patients.

4.2.7.2 Randomized period

DCR by investigator will be analysed using a logistic regression stratified by race (Chinese/Asian vs. Non-Chinese/Asian vs. Non-Asian), WHO PS (0 vs. 1), and method used for tissue testing (central vs. local) The results of the analysis will be presented in terms of an odds ratio together with its associated 95% profile likelihood CI and two-sided p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model that contains the covariates defined above).

The analysis of DCR will take place at the time of the primary PFS analysis only.

4.2.8 Depth of response (Change in TL tumour size)

4.2.8.1 Safety run-in period

The best absolute change and best percentage change in TL tumour size from baseline will be summarized descriptively for each treatment cohort, and will be presented in a waterfall plot for all safety run-in patients.

4.2.8.2 Randomized period

Depth of response (i.e. tumour shrinkage / change in tumour size) will be examined by presenting the proportion of patients who achieve >30%, >50% and >75% reduction from baseline in target lesion tumour size.

The best absolute change and best percentage change in TL tumour size from baseline will be summarized descriptively and presented by randomized treatment group. The best percentage change in TL tumour size will be presented graphically using waterfall plots, with each patient's best percentage change in tumour size represented as a separate bar and the bars ordered from the largest increase to the largest decrease. Reference lines at the +20% and –30% change in tumour size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively.

The effect of osimertinib plus chemotherapy treatment on best percentage change in TL tumour size will be estimated from an analysis of covariance (ANCOVA) model with covariates for race (Chinese/Asian vs. Non-Chinese/Asian vs. Non-Asian), WHO PS (0 vs. 1), and method used for tissue testing (central vs. local), baseline tumour size and time from baseline scan to randomization. The number of patients, unadjusted mean, and least squares means for each randomized treatment group will be presented, together with the difference in least squares means, 95% CI and corresponding p-value.

4.2.9 Subgroup analyses for primary PFS – Randomized period

In addition to the analysis of PFS described in section 4 2.3 2, the following subgroup analyses will be conducted by comparing PFS between treatments (i.e., using a Cox-Proportional Hazards Model) in the following groups:

- Gender (Male, Female)
- Race (Chinese/Asian, non Chinese/Asian, non-Asian)
- Method used for tissue testing (central vs. local)
- Age at screening (< 65 years, ≥ 65)
- Smoking history
- EGFR mutation type (Exon 19 Deletion or L858R)
- EGFR by central ctDNA cobas test (Positive, Negative, Missing)
- EGFR mutations by central cobas tissue test (Positive, Negative, Missing)
- WHO performance status (0, 1)
- CNS status at baseline (yes, no)
- Central confirmation of EGFR mutation (centrally confirmed tissue or ctDNA EGFR positive result, no central confirmation)

To classify patients' CNS status at baseline (Yes, No) the electronic CRF (eCRF) will be used. Patients who have CNS metastases at baseline, or history of CNS metastases, will have a CNS status at baseline of Yes. Otherwise, patients will have a CNS status at baseline of No.

For each subgroup listed above, the HR and 95% CI will be calculated from a single Cox proportional hazards model that contains a term for treatment, the subgroup covariate of interest and the treatment by subgroup interaction term. The treatment effect HR will be obtained for

each level of the subgroup from this model. The Cox models will be fitted using SAS® PROC PHREG with the Efron method for handling ties.

These HRs and associated two-sided 95% CIs will be summarized and presented on a forest plot, along with the results of the overall primary analysis. In addition, a Cox proportional hazards model that contains a term for treatment will be fitted and the treatment effect HR and the two-sided 95% CIs will also be presented on the forest plot.

If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 20 events per level in a subgroup), the relationship between that subgroup and PFS will not be analysed via a Cox proportional model. In this case, only descriptive summaries will be provided.

No adjustment to the significance level for testing will be made since the subgroup analysis may only be supportive of the primary analysis of PFS.

4.2.10 Time from randomization to second progression or death (PFS2) -Randomized period

Analysis of Time from randomization to second progression or death (PFS2) will take place at the time of the primary PFS analysis only.

PFS2 will be analysed using the same method as the analysis of PFS with the exception that the sensitivity, subgroup and exploratory analyses will not be performed. Medians, quantiles, and a Kaplan-Meier plot will be provided to support the analysis.

The number and percentage of patients experiencing a PFS2 event (death or second progression) and the number and percentage of patients prematurely censored will be summarized by randomized treatment arm.

4.2.11 Time to first and second subsequent therapy or death (TFST and TSST) - Randomized period

The time to first subsequent treatment (TFST) and time to second subsequent treatment (TSST) will be analyzed in the FAS using a log rank test stratified by race (Chinese/Asian vs. Non-Chinese/Asian vs. Non-Asian), WHO PS (0 vs. 1), and method used for tissue testing (central vs. local). The HR and CI will be obtained using the same method for primary PFS.

The number (%) of patients received a subsequent anti-cancer therapy (first and second) and the number (%) of patients censored will be summarized by randomized treatment arm.

In addition, medians, quantiles, and Kaplan-Meier plots of TFST/TSST will be presented.

4.2.12 Patient reported outcomes (PROs) – Randomized Period

The analysis population for PRO data will be the FAS.

QLQ-C30 and QLQ-LC13

Mixed models repeated measures (MMRM) of change from baseline in primary PRO symptoms & function scores

Change from baseline in the primary PRO scores for cough, dyspnea, chest pain, fatigue, appetite loss, physical function and global health status/QoL will be analyzed separately for each treatment comparison using a mixed model for repeated measures (MMRM) analysis with use of all data from baseline up to PD or 19 months. The analysis will compare the average treatment effect from the point of randomization until PD or 19 months (whichever is earlier), excluding visits with excessive missing data (defined as more than 75% missing data).

It is acknowledged that patients will discontinue treatment at different timepoints during the study and that this is an important time with regards to PRO data collection. To account for this and in order to include the discontinuation and follow up visits, all PRO assessments will be mapped to week *z* per the analysis visit window specified below. Analysis visits will be mapped for PRO assessments occurring before PD or within the nineteen months (less or equal to 19×30.4375 study day) , whichever is earlier, of randomization

If there are two or more values potentially allocated to the same scheduled assessment, the post baseline assessment closest to the scheduled assessment date will be included in the summaries and in the MMRM.

The MMRM model will include subject, treatment, assessment (after mapping based on the study day window as in Table 21 and Table 22) and treatment by assessment interaction as explanatory variables, with the baseline PRO score as a covariate along with the baseline PRO score by assessment interaction. Treatment, assessment and treatment-by-assessment interaction will be fixed effects in the model; subject will be included as a random effect. Restricted maximum likelihood (REML) estimation will be used. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over assessments giving each visit equal weight.

For this overall treatment comparison, adjusted mean estimates per treatment group and corresponding 95% CIs will be presented along with an estimate of the treatment difference, 95% CI. No p-values will be presented. The treatment-by-assessment interaction will remain in the model regardless of significance.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. The following provides sample code for implementing the MMRM analysis:

```
proc mixed data=PRO method = reml;
        class TRT(ref="SoC") WEEK SUBJECT;
        model PROSC = TRT WEEK TRT*WEEK PROBL PROBL*WEEK /s ddfm=kr;
        repeated WEEK / type=UN subject=SUBJECT;
        lsmeans TRT / at means diff alpha=0.05 cl;
        run;
```

where TRT is the randomized treatment, WEEK is the assessment, PROSC is the change from baseline in the PRO score, and PROBL is the baseline PRO score.

For the estimation of TRT*WEEK means an additional model will be run using all visits and the following lsmeans statement:

lsmeans TRT*WEEK / slice=WEEK diff alpha=0.05 cl;

If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, and autoregressive. If there are still issues with the fit of the model or estimation of the treatment effects, SUBJECT will be treated as a fixed effect.

Change from baseline, time to deterioration (TTD), and compliance

Descriptive statistics and line plots will be produced for the primary PRO symptom scores by scheduled assessments along with the change from baseline. These will also be reported for the other EORTC QLQ-C30 and EORTC QLQ-LC13 reported i ems and scales.

For each PRO subscale or single item, Kaplan Meier plot will be presented for TTD by treatment group. Summaries of the number and percentage of patients who have an event as well as who were censored will be provided along with K-M medians for each treatment. The hazard ratio and it's corresponding 95% CI for each subscales will be analysed using the same method as primary PFS.

Time to the definitive deteriation (TTDD) will be analyzed as the same as TTD, as sensitivity analysis for each PRO subscale or single item.

The assessments used in the MMRM model ("WEEK" in the proc mixed procedure above) for the change from baseline of QLQ-LC13 and QLQ-C30 subscales will be mapped as Table 21and Table 22 (study day will be calculated in relation to the date of first study dose administration. In case of first study dose administration date are not available, study day will be calculated in relation to the date of randomization):

- EORTC QLQ-LC13 data will be collected weekly starting at C1D1 until day 57. From day 64 every 3 weeks.

Table 21 EORT QLQ-LC13 and PRO-CTCAE assessment window mapping

Clinical Visit Name	Assessment in MMRM model	Scheduled Assessment Study Day	Assessment Window (study Day)
Cycle 1 Day 1	Week 1 (baseline)	1	(-1, 2)
Cycle 1 Day 8	Week 2	8	(3, 11)
Cycle 1 Day 15	Week 3	15	(12, 18)
Cycle 2 Day 1	Week 4	22	(19, 25)
Cycle 2 Day 8	Week 5	29	(26, 32)

Clinical Visit Name	Assessment in MMRM model	Scheduled Assessment Study Day	Assessment Window (study Day)
Cycle 2 Day 15	Week 6	36	(33, 39)
Cycle 3 Day 1	Week 7	43	(40, 46)
Cycle 3 Day 8	Week 8	50	(47, 53)
Cycle 3 Day 15	Week 9	57	(54, 60)
Cycle 4 Day 1	Week 10	64	(61, 74)
Cycle 5 Day 1	Week 13	85	(75, 95)
Cycle 6 Day 1	Week 16	106	(96, 116)
Cycle 7 Day 1	Week 19	127	(117, 137)
Cycle *x* Day 1 thereafter (*x*>=5)	Week 21, then every 3 weeks thereafter before EOT	Day (*x* 1)*21+1	[(*x*-1)*21+1-10, (*x*-1)*21+1+10]
EOT, and follow up visits until PD or 19 months (week 82) after randomization*	Week *y*	NA	Use the same window (weekly or every 3 weeks) of the latest assessment before early discontinuation/ EOT

*Continue mapping a visit after week 82 (every 3 weeks) until PD for compliance summary, but only includes records prior to PD or prior or on week 82 in MMRM model.

- EORTC QLQ-C30 data will be collected starting at C1D1, day 22, day 43, and every 6 weeks starting from day 64.

Table 22 EORT QLQ-C30, PGIS and EQ-5D-5L assessment window mapping

Clinical Visit Name	Assessment in MMRM model	Scheduled Assessment Study Day	Visit Window
Cycle 1 Day 1	Week 1 (baseline)	1	(-1, 2)
Cycle 2 Day 1	Week 4	22	(3, 32)
Cycle 3 Day 1	Week 7	43	(33, 53)
Cycle 4 Day 1	Week 10	64	(54, 84)
Cycle 6 Day 1	Week 16	106	(85, 126)
Cycle 8 Day 1	Week 22	148	(127, 168)
Cycle 10 Day 1	Week 28	190	(169, 210)
Cycle *x* Day 1 thereafter (*x*>=12)	Week 34,	Day (*x*-1)*21+1	[(*x*-1)*21+1-21, (*x*-1)*21+1+20]

Clinical Visit Name	Assessment in MMRM model	Scheduled Assessment Study Day	Visit Window
	every 6 weeks thereafter before EOT		
EOT, and follow up visits until PD or 19 months after randomization*	Week *y*	NA	Use the same window (every 3 weeks or every 6 weeks) of the latest assessment before early discontinuation/EOT

*Continue mapping a visit after week 82 (every 6 weeks) until PD for compliance summary, but only includes records prior to PD or prior or on week 82 in MMRM model.

If there is more than one value per patient within a time window then the closest value to the scheduled visit date will be summarized and analyzed, or the earlier, in the event the values are equidistant from the nominal visit date.

PRO data will be collected every 8 weeks post progression until the second disease progression or DCO for primary PFS analysis, whichever is sooner.

QLQ-LC13 and QLQ-C30 data post disease progression will not be included in the MMRM model, but will be summarized as 8 weeks post PD, 16 weeks post PD, every 8 weeks post PD afterwards until the second disease progression r DCO for primary PFS analysis, whichever is sooner. The later PRO assessment will be selected for summary if multiple records within the same every 8 weeks window.

PRO-CTCAE, PGIS, and EQ-5D-5L

The PRO-CTCAE and PGIS data will be presented using summaries and descriptive statistics.

EQ-5D index score along with the change from baseline at each scheduled visit will be presented by randomized treatment arm.

PRO compliance rate will be summarized by scheduled assessment time point.

To support submissions to payers, additional analyses may be undertaken and these will be outlined in a separate Payer Analysis Plan (PAP) for PROs.

4.2.13 CNS Efficacy Analysis – Exploratory Analysis for Randomized Period

4.2.13.1 Patients in cFAS

The cFAS analysis set will be used for the CNS efficacy exploratory analysis. The main exploratory endpoints for the CNS analysis is CNS PFS by CNS BICR.

CNS PFS by CNS BICR assessment

A summary of the number and percentage of patients experiencing a BICR CNS PFS event, and the type of event (CNS progression or death) will be provided, along with medians (and 95% CI) per treatment arm.

BICR CNS PFS will be analysed using a K-M analysis. Median CNS PFS (calculated from the K-M plot, with 95% CIs), and the percentage of patients alive and CNS progression free at 6-monthly intervals will be evaluated until there are no data available at that timepoint. The analysis method will be the same as the primary PFS analysis (described in Section 4.2.3.2). Stratification factors will be collapsed if there are small number of events within a stratum (i.e. <5) . Sensitivity analysis and subgroup analysis will not be performed for CNS PFS.

CNS ORR by CNS BICR assessment

A summary of the best objective CNS response will be presented by randomized treatment group. CNS ORR by BICR will be analysed using a logistic regression, same covariates in the model as the analysis of ORR by Investigator. If there are small number responders within a stratum (i.e. <5), unstratified CMH test will be used to analyse the data.

CNS BICR assessment –reason for CNS progression

The number of patients with and without CNS metastases at RECIST progression will be tabulated by CNS disease status at baseline (present/absence).

A summary table will be included to present the number/percentage of patients with RECIST progression by whether there was documented progression in the CNS. The summary table will indicate whether the CNS progression w s in a target, non-target, or a new lesion. The table will present the data by randomized treatment arm.

For patients with CNS progres ion due to a new CNS lesion, a summary will be provided to show the extent of CNS disease at baseline, and the response of CNS target/non-target lesions at the time of the appearance of the CNS new lesion.

CNS DoR and CNS DCR by CNS BICR assessment;

Descriptive data will be provided for the CNS DoR. CNS DCR will be analysed use the same method and stratifications as DCR. If there are small number responders within a stratum (i.e. <5), unstratified CMH test will be used to analyse the data.

Best percentage change in CNS tumour size (target lesion) by CNS BICR assessment

Up to 5 measurable lesions in the brain will be selected at baseline. The best change and best percentage change in CNS TL tumour size will be summarized descriptively, and a waterfall plot will be provided by randomized treatment arm.

4.2.13.2 CSN PFS for FAS Patients

CNS PFS by CNS BICR will be analyzed for patients with/without history of CNS metastases at baseline when the sufficient events (≥ 20 events) are observed for a meaningful analysis, using the same methodology as for the primary PFS by investigator.

Summaries for presence or absence of CNS lesions at RECIST progression (by BICR) will be provided for patients with/without (history of) CNS metastases at baseline..

4.2.14 Biomarker analysis – Exploratory Analysis for Randomized Period

Biomarker analysis will be discussed and reported in separate documents.

4.2.15 Health care resource use

Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) will be provided for each randomized treatmen arm or treatment cohort for the safety run-in on the different types of hospital admissions, the length of stay of people admitted in to hospital for at least one overnight stay and length of stay of people admitted to intensive care / high dependency units, as well as the primary sign or symptom the patient presents with.

To support submissions to payers, additional analyses may be undertaken and these will be outlined in a separate Payer Analysis Plan.

4.2.16 Data cut-offs

At analysis for PFS or OS, a survival swe p should be performed within 7 days after each DCO to ensure that complete overall surviva data is collected, updating the survive and death raw data to ensure the most accurate OS analy es possible.

4.3 Safety and Demographics

4.3.1 Safety

A Safety Review Committee for the safety run-in will be utilized for this study.

When at least 12 patients in each cohort have either received ≥ 3 cycles of osimertinib with chemotherapy (osimertinib, cisplatin or carboplatin, and pemetrexed) or have discontinued study treatment due to unacceptable toxicity, a Safety Review Committee will convene. A formal database lock will be required to facilitate this data review. All data, including, safety, tolerability, and available PK data from all patients, will be reviewed.

Full details of the Safety Review Committee procedures and processes can be found in the Safety Review Committee Charter.

Independent Data Monitoring Committee (IDMC) for the randomized period will be utilized for this study.

Oversight of safety and tolerability of the randomized period of the study will be provided by an IDMC which will be comprised of fully independent members. The IDMC will meet

periodically to review safety data and will make recommendations to continue, amend, or stop the study based on findings. Serious AEs, AEs, and other safety data will be reviewed, and individual and aggregated safety data will be evaluated by the IDMC.

Full details of the IDMC procedure and processes can be found in the IDMC Charter.

4.3.1.1 General considerations for safety

For summaries of vital signs, laboratory data, and ECG assessments will be assigned to calculated visit windows (using study day). Time windows is defined for any presentations that summarise values by visit. The following conventions will apply:

- The time windows will be exhaustive so that data recorded at any time point has the potential to be summarized. Inclusion within the time window will be based on the actual date and not the intended date of the visit.
- All unscheduled visit data have the potential to be included in the summaries.
- The analysis window for the visits following baseline will follow the visit window specified in CSP SoA table be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two c nsecutive visits then the upper limit will be taken as the midpoint value minus 1 day.
- Analysis visit window for vital signs laboratory data, ECGs (exclude Japan)

Visit Name	**Normal Visit Study Day**	**Visit Window (Study Day)**
Cycle 1 Day 8	Day 8	(2, 11)
Cycle 1 Day 15	Day 15	(12, 18)
Cycle 2 Day 1	Day 22	(19, 32)
Cycle 3 Day 1	Day 43	(33, 53)
Cycle 4 Day 1	Day 64	(54, 74)
Cycle 5 Day 1	Day 85	(75, 95)
Cycle 6 Day 1	Day 106	(96, 116)
Cycle 7 Day 1	Day 127	(117, 137)
Cycle *x* Day 1 thereafter (*x*>=3)	Day (*x*-1)*21+1	[(*x*-1)*21+1-10, (*x*-1)*21+1+10]

- Analysis visit window for vital signs, laboratory data, ECGs (Japan only)

Visit Name	**Normal Visit Study Day**	**Visit Window (Study Day)**
Cycle 1 Day 8	Day 8	(2, 11)
Cycle 1 Day 15	Day 15	(12, 18)

Visit Name	Normal Visit Study Day	Visit Window (Study Day)
Cycle 2 Day 1	Day 22	(19, 25)
Cycle 2 Day 8	Day 29	(26, 36)
Cycle 3 Day 1	Day 43	(37, 53)
Cycle 4 Day 1	Day 64	(54, 74)
Cycle 5 Day 1	Day 85	(75, 95)
Cycle 6 Day 1	Day 106	(96, 116)
Cycle 7 Day 1	Day 127	(117, 137)
Cycle *x* Day 1 thereafter (*x*>=3)	Day (*x*-1)*21+1	[(*x*-1)*21+1-10, (*x*-1)*21+1+10]

- For visit based summaries:
  - If there is more than one value per patient within a time window then the closest value to the scheduled visit date will be summarized, or the earlier, in the event the values are equidistant from the nominal visit date. If duplicate observations are recorded on the same day the average of the two values will be selected for analysis at that visit. Any repeat or additional assessments performed will be included in the individual patient data listings. The listings will highlight the value for the patient that contributed to the summary table, wherever feasible. Note: in summaries of extreme values all post baseline values collected are used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date.
  - To prevent very large tables or plots being produced that contain many cells with meaningless data, for each tr atment group, visit data will only be summarized if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.
  - LVEF data will be summarized based on eCRF collected visits.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For summaries at a patient level, all values will be included, regardless of whether they appear in a corresponding visit based summary, when deriving a patient level statistic such as a maximum.
- Baseline for safety assessments will generally be the last value obtained prior to the first dose of study medication (for PRO baseline see section 4.1.3), except the baseline of ECG will be the average of triple ECGs at the last timepoint prior to the first dose. Alternatively, if two visits are equally eligible to assess patient status at baseline (e.g., screening and baseline assessments both on the same date prior to first dose with no washout or other intervention in

the screening period), the average will be taken as a baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline and the worst value would be taken for post-baseline records as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline. Where safety data are summarized over time, study day will be calculated in relation to date of first treatment.

4.3.1.2 Adverse events (AEs)

All AEs, both in terms of Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) and CTCAE grade, will be listed and summarized descriptively by count (n) and percentage (%) for each treatment arm. Latest available MedDRA version will be used for coding. Missing coding terms should be listed and summarized as "Not coded".

Any AE occurring before start of study treatment (i.e. before Dose Day 1) will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'.

Any AE occurring after the first dose and within 28 days of discontinuation of IP (i.e. the last dose of study treatment) but prior to or on the start date of a subsequent anti-cancer treatment will be included in the AE summaries. If the start date of the first subsequent anti-cancer therapy is completely missing or has missing day and month, AEs onset within the last dose + 28 days will be included in AE summaries. If the tart date of the first subsequent anti-cancer therapy is partial with missing day, then month and year of the date will be compared to the AE onset date. AEs onset within the last dose + 28 days and on or before the month/year of the subsequent anti-cancer therapy date will be included in AE summaries. Any AE in this period that occur after a patient has received further therapy for cancer (following discontinuation of IP) will be flagged in the data listings.

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved), investigator's assessment of CTCAE grade and relationship to study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from on the episode level summaries).

Summary information (the number and percent of patients by treatment) by System organ class (SOC) and preferred term (PT) will be tabulated for:

- All AEs
- All AEs causally related to study medication
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related to treatment
- AEs with outcome of death
- AEs with outcome of death causally related to treatment

- AEs leading to dose reduction
- AEs leading to dose interruption
- AEs leading to dose modification (dose reduction or dose interruption)
- All SAEs
- All SAEs causally related to study medication
- AEs leading to discontinuation of treatment
- AEs leading to discontinuation of treatment, causally related to treatment
- OAEs
- OAEs causally related to treatment

An overall summary of the number and percentage of patients in each category will be presented, as well as an overall summary of the number of events in each category. In addition, a truncated AE table of most common AEs, showing all events that occur in at least 5% of patients overall will be summarized by preferred term, by decreasing frequency. This cut-off may be modified after review of the data.

AEs will be assigned CTCAE grades and summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, SOC, PT and actual treatment group. Fluctuations observed in CTCAE grades during tudy will be listed (where collected).

In addition, AEs with outcome of death, SAEs, AEs leading to discontinuation of each study drug, AEs causally related to each study drug, OAEs, and AESIs will be listed.

A separate summary of AEs occurring more than 28 days after discontinuation of study treatment or on/after the start of subs quent anti cancer treatment (where reported) as well as those occurring prior to treatment will be produced. These events will not be included in AE summaries.

4.3.1.3 Adverse events of special interest (AESI)

Preferred terms used to identify AESI will be listed before DBL and documented in the Trial Master File. Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI grouping. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

An overall AESI summary of the above-mentioned grouped AE categories will include number (%) of patients who have:

- At least one AESI presented by event outcome
- At least one AESI causally related to study medication
- At least one AESI leading to discontinuation of study medication

- At least one AESIs leading to dose interruption of study medication
- At least one AESIs leading to dose reduction of study medication
- At least one AESIs leading to dose modification (dose reduction or dose interruption) of study medication
- At least one serious AESI
- At least one AESI with CTCAE grade 3 or higher

A summary of total duration (days) of AESI will be provided for events which have an end date and this may be supported by summaries of ongoing AESIs at death and, separately, at data cut-off.

Summary tables of AEs of special interest will be produced. The number (%) of patients experiencing any of the specified terms will be presented overall and by maximum CTCAE grade.

Additional summaries of time to onset of first AE for each grouped term and each preferred term within it; time to onset of first CTCAE grade three or higher and duration of AEs of special interest will be produced.

In addition, further summary tables from the AEs section listed above will be repeated for grouped AEs of special interest.

The number (%) of patients with AEs with a body system of Infections and Infestations occurring with concomitant low leukocyte/neutrophil counts (below the normal range) will be summarized. Only treatment-emergent events with an onset after the date of a treatment emergent low leukocyte/neutrophil count value, and before the date when the leukocyte/neutrophil value returns to normal, will be presented.

The above summary of AEs with a body system of Infections and Infestations will also be repeated for concomitant low neutrophil counts.

The number (%) of patients with AEs of bleeding occurring with concomitant low platelet counts (below the normal range) will be summarized. Bleeding adverse events will be identified using the "Hemorrhages" narrow MedDRA standardized MedDRA query (SMQ). Only treatment-emergent bleeding events with an onset after the date of a treatment-emergent low platelet count value, and before the date when the platelet value returns to normal, will be presented.

4.3.1.4 Summary of long-term tolerability

Life table plot, prevalence plot , and cumulative incidence plot for each of the AESI grouped terms and any other events considered important after review of the safety data, provided there are ≥ 10 events across both arms, will also be produced.

Kaplan Meier curves for time to onset of each AESI grouped term and any other events considered important after review of the safety data, provided there are ≥ 10 events, will be provided. Time to onset of AE is defined as date of first AE – date of first dose+1 for those

patients with an AE, and min ([date of last dose + min(28-day follow-up period, start date of subsequent anti-cancer therapy)], OS date, DCO) – date of first dose + 1 for those without an AESI (i.e. censored patients).

A prevalence plot provides information on the extent to which the events may be an ongoing burden to patients. The prevalence at time t after first dose of study treatment is calculated as the number of patients experiencing the event divided by the number of patients receiving study treatment or in safety follow-up at time t; generally, t is categorized by each day after dosing. The prevalence is plotted over time split by treatment arm. Multiple occurrences of the same event are considered for each patient but a patient is only counted in the numerator whilst they are experiencing one of the occurrences of the event. These plots will only be produced for AESIs that have ≥10 events.

A life table plot can be used to describe the time to onset of the event and specifically when patients are at most risk of first experiencing the event. The hazard, or in other words, the probability of having an AE in a specified time period (e.g. 0-1 months, 1-3 months, 3-6 months, etc.) given that the patient reaches that time period without having an event is plotted for each time period split by treatment. These plots will only be produced for AESIs that have ≥10 events.

A cumulative incidence plot is a plot of the raw cumulative incidence and cumulative incidence function over time with the treatment groups presented on separate plots. The raw cumulative incidence is the actual probability that a patient will have experienced their first occurrence of the event by a given time point. The cumulative incidence function estimates the cumulative incidence if the data cut-off had not been imposed and all patients had completed 28-day follow-up period. These plots will only be produced for AESIs that have ≥10 events.

4.3.1.5 Death

A summary of deaths will be pro ided with number and percentage of patients, categorized as:

- Related to disease under investigation only,
- AE outcome = death only,
- Both related to disease under investigation and with AE outcome=death,
- AE with outcome of death only (AE start date falling after 28 days follow-up period, or after the start of new subsequent anti-cancer therapy),
- Other death (not captured in above categories)
- Patients with unknown reason for death.

A corresponding listing will be produced.

4.3.1.6 Laboratory evaluations

All laboratory data recorded in the eCRF will be listed.

If any additional analytes to those in Table 14 are also recorded then these will be listed only. All values will be classified as low (below range), normal (within range) and high (above range) based on project-specific reference ranges. As applicable, values will be converted to standard units and will be graded using the latest versions of CTCAE.

For clinical chemistry and hematology, shift tables will present movements from baseline to maximum or minimum (as applicable) according to reference range classification and CTCAE grade changes from baseline to the maximum grade on treatment will be provided. Corresponding shift tables ("Negative", "Trace", "Positive", "0", "+", "++", "+++", ">+++") will be produced for urinalysis.

A patient list for patients with potential Drug Induced Liver Injury (i.e. ALT or AST ≥ 3×ULN and TBIL ≥ 2×ULN, ALT or AST ≥ 5×ULN will be provided.

Liver biochemistry tests (ALT, AST, TBIL, ALP, GGT) results over time will also be presented graphically for patients with potential Drug Induced Liver Injury

Plots of both the maximum post-baseline alanine aminotransferase (ALT) and aspartate aminotransferase (AST) versus the maximum post-baseline total bilirubin, expressed as multiples of their upper limit of reference range will be produced.

Box plots of absolute values and change from baseline for all hematology and clinical chemistry parameters will also be presented.

Pregnancy testing results will be presented in a listing.

4.3.1.7 Vital signs (pulse and BP) and weight

Absolute values and change from baseline at each scheduled visit for pulse, BP and weight will be summarized by randomiz d treatment group or by treatment cohort for the safety run-in.

4.3.1.8 Physical examination

Abnormalities identified from physical examination will be summarized and listed.

4.3.1.9 ECG

All ECG data received will be presented in data listings. The average of triple ECG results at the same timepoint will be used for the summaries. In case of missing one or two assessment(s) of triple ECGs, the average of available results will be used.

ECG summaries will be presented for patients in the safety analysis set. The following ECG parameters will be summarized (absolute values and change from baseline) by visit: QTcF, RR variability, PR interval, QRS complex, and QT interval.

Box plots for observed ECG parameters and change from baseline in ECG parameters over time will be presented.

Shift plots of the value corresponding to the maximum absolute change from baseline versus the baseline value for QTcF, with reference lines for 450 msec, ±30 msec and ± 60 msec change, will be presented.

QTc outliers are defined as QTcF values following dosing that are greater than 450 msec or are increases from baseline greater than 30 msec.

QTcF outliers will be highlighted in the data listings and summarized using the following categories:

- Values > 450 msec, > 480msec, > 500 msec
- Increase from baseline of > 30 msec, Increase from baseline of > 60 msec, Increase from baseline of > 90 msec
- Values > 450 msec and increases of > 30 msec, Values > 500 msec and increases of > 60 msec

The number and percentage of patients who meet the ECG outlier criteria at any assessment post-date of first dose will be summarized.

4.3.1.10 Left ventricular ejection fraction

Plots of absolute LVEF values and change from baseline in LVEF values over time will be presented.

LVEF outliers for patients who have both baseline value and at least one post baseline assessment are defined as LVEF values (including unscheduled assessment) following dosing that are

- ≥10 percentage points decrease from baseline and < 50%, or
- ≥15 percentage points decrease from baseline and ≥ 50%.

The number of patients with the following LVEF values at each post-baseline scheduled LVEF visit, and the maximum post-baseline change will be displayed:

- LVEF increase
  - ≥ 30%
  - ≥ 20 - < 30%
  - ≥ 10 - < 20%
- LVEF change < 10%
- LVEF decrease
  - ≥ 10 - < 20% and absolute value < 50%
  - ≥ 10 - < 20% and absolute value ≥ 50%
  - ≥ 20 - < 30% and absolute value < 50%

- ≥ 20 - < 30% and absolute value ≥ 50%
- ≥ 30% and absolute value < 50%
- ≥ 30% and absolute value ≥ 50%

For the maximum change, patients with a maximum increase ≥ 10% and a maximum decrease < 10% will be summarized under their maximum increase, and patients with a maximum decrease ≥ 10% and a maximum increase < 10% will be summarized under their maximum decrease.

4.3.1.11 WHO performance status

WHO performance status will be listed and summarized as frequency counts by visit and by treatment group or treatment cohort for the safety run-in.

4.3.2 Demographics and baseline characteristics

Safety run-in demographics and baseline characteristics summaries will be based on SAF, and based on FAS for randomized period.

The following will be summarized for all patients by randomized treatment group or by treatment cohort for the safety run-in period:

- Patient disposition
- Important protocol deviations
- Inclusion in analysis sets
- Demographics (age, age group (<50, ≥50 and < 65 years, ≥ 65 and <75, ≥75), sex, race and ethnicity)
- Patient characteristics at baseline (height, weight)
- Patient recruitment by country and centre
- EGFR mutation testing method (Central test, local test) and mutation type (EGFR Exon 19 Deletion, EGFR Exon 21 L858R) at randomization
- EGFR mutations by central cobas tissue test (EGFR Exon 19 Deletion, EGFR Exon 21 L858R)
- EGFR mutations by central ctDNA cobas test (EGFR Exon 19 Deletion, EGFR Exon 21 L858R)
- Central cobas EGFR mutation status (EGFRm detected, EGFRm not detected, EGFRm unknown) for patients randomized by local EGFRm
- WHO performance status (0, 1) at baseline
- CNS status at baseline (yes, no)
- Previous radiotherapy
- Palliative concomitant radiotherapy
- Previous chemotherapy prior to this study

- Disease characteristics at baseline (primary tumour location, histology type, tumour grade and overall disease classification)
- Extent of disease upon entry to study
- Medical history (past and current)
- Relevant surgical history
- Concomitant medications
- Post-treatment cancer therapy
- Post-treatment radiotherapy
- Nicotine use, categorized (never, current, former)

Disposition summaries will include the number and percentage of patients:

- Enrolled (informed consent received)
- Randomized (randomization period)
- Patients not randomized (randomization period)
- Included in each analysis set.
- Patients ongoing study treatment at the data cut-off
- Patients discontinued the treatment still on 28 Day follow-up, PFS follow-up, and OS follow-up visits

In addition, the number and percentage of patients who discontinued treatment and who discontinued the study, including a breakdown of the primary reason for discontinuation will be presented.

Medical history and relevant surgical history will be coded using the latest available MedDRA version. All medical history (past and current) will be listed and the number and percentage of patients with any medical history will be summarized by system organ class (SOC) and preferred term (PT). Relevant surgical history will be summarized similarly. All surgical history will be listed.

All important protocol deviations will be listed and summarized. All protocol deviations will be defined by the study team before database lock.

4.3.3 Concomitant and other treatments

Prior medications, concomitant and post treatment medications summaries will be provided for safety run-in period and randomized period, based on SAF for safety run-in and FAS for randomized period.

Information on any treatment within the four weeks prior to initiation of study drug and all concomitant treatments given up to 28 days after discontinuation of study treatment, or objective disease progression (whichever is later), with reasons for the treatment, will be recorded in the eCRF. Thereafter, only subsequent regimens of anti-cancer therapy will be recorded in eCRF.

Other anti-cancer therapies, investigational agents, and radiotherapy should not be given while the patient is on study drug.

Medications received prior to, concomitantly, or post-treatment will be coded using the AstraZeneca Drug Dictionary Anatomical Therapeutic Chemical (ATC) Classification codes. Concomitant medications will be summarized by ATC classification codes.

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates will be imputed as detailed in Section 4.1.4.1.

Prior medications, concomitant and post treatment medications are defined based on imputed start and stop dates as follows:

- Prior medications are those taken prior to study treatment with a stop date prior to the first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment (and could have started prior to or during treatment).
- Post-treatment medications are those with a start date after the last dose date of study treatment.

In addition, all post-treatment anti-cancer medications and surgical procedures will be summarized by randomized treatment group or by treatment cohort for the safety run-in.

The following summaries will be produced:

- Summary of prior medications
- Summary of concomitant medications
- Summary of Post study treatment cancer therapies

All concomitant and other treatment data will be listed.

Missing coding terms should be listed and summarized as "Not coded".

4.3.4 Exposure

Exposure will be summarized by study drug and overall of all three study drugs on the safety analysis set. The following summaries will be produced:

- Summary of duration of exposure (including total exposure and actual exposure) of study treatment for osimertinib
- Summary of total exposure for chemotherapy
- Number of cycles chemotherapy received
- Summary of dosing interruptions and reductions for Osimertinib
- Summary of duration of dosing interruptions for Osimertinib

- Summary of dosing interruptions, reductions, and delays for chemotherapy (for cisplatin, carboplatin, and pemetrexed respectively)
- Summary of total actual dose (mg), total planned dose (mg), relative dose intensity (RDI) (%), and percentage intended dose (PID) (%) for Osimertinib and Pemetrexed

The dosing administration of osimertinib, cisplatin, carboplatin, and pemetrexed will be listed by patient.

4.3.5 Pharmacokinetic data

Plasma concentrations of osimertinib and metabolite AZ5104 will be summarized by nominal sample time using standard summary statistics for PK concentrations (geometric mean, geometric coefficient of variation, geometric mean ± standard deviation, arithmetic mean, standard deviation, minimum, maximum and n) within each treatment cohort. PK parameters will also be summarized within each treatment cohort. All plasma concentrations and PK parameters will be listed regardless of whether they're excluded from summary statistics due to deviation (e.g. as a result of dose interruption, reduction or missing the dose before PK sample collection, or sampling time deviation, etc).

Outputs will display the data as described in the 'Reporting formats' tab of the PK Order Form spreadsheet.

4.3.6 Impact of COVID-19

Depending on the extent of any impact, summaries of data relating to impact of COVID-19 on study conduct (in particular missed visits, delayed or discontinued IP, and other protocol deviations) may be generated by treatment group, including:

- Disposition (discontinued IP due to COVID-19 and withdrew study due to COVID-19)
- Deviations (overall deviations plus if due to COVID-19 and not due to COVID-19)
- Summary of COVID-19 disruption (visit impact, drug impacted)
- Listing for patients affected by the COVID-19 pandemic
- Listing for patients with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic.
- Protocol deviations
- Medical history within section 4.3.2
- Baseline characteristics
- Supplementary Analysis for PFS /OS - Accounting for COVID-19 deaths (e.g. In the case that there is a sufficient number of patients with a confirmed or suspected COVID-19 death event (either at least five patients and/or at least 2% of the patient population), a supplementary analysis will be conducted to assess the potential impact of COVID-19 deaths on PFS /OS)

- COVID-19 Related AEs within section 4.3.1.2.

5. IDMC AND INTERIM ANALYSES

5.1 Safety run-in period

A Safety Review Committee for the safety run-in period will be utilized for this study.

Data for IDMC review will not combine data from the safety run-in and randomized period to assess safety. The data will be reviewed separately.

No interim analysis will be performed during the safety run-in period.

5.2 Randomized period

IDMC

Oversight of safety and tolerability of the randomized period of the study will be provided by an Independent Data Monitoring Committee (IDMC), with fully independent members. The IDMC will convene at the beginning of the randomized period, but will have access to up to date data from patients ongoing in the safety run-in period as well as the randomized period to inform their recommendations.

The IDMC will meet after data are available from approximately 60 patients across both randomized treatment arms with at least 28 days of follow-up. Thereafter, the IDMC will meet after data are available from approximately 150 patients across both treatment arms with at least 28 days of follow-up, after 300 patien s across both treatment arms with at least 28 days of follow-up, after completion of recruitment, and approximately every 6 months until the primary PFS analysis DCO.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca and are free from conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will include the recommendation and any potential protocol amendments, and will not contain any unblinding information.

The final decision to modify or stop the study will sit with the sponsor. The sponsor or IDMC may call additional meetings if at any time there is concern about the safety of the study.

Full details of the IDMC procedures and processes can be found in the IDMC Charter.

Interim analysis for futility

A futility analysis will be conducted prior to the primary analysis of PFS for the randomized period to assess for a potential lack of efficacy in the osimertinib with chemotherapy arm when compared with osimertinib monotherapy. An IDMC will review the efficacy and safety data of the randomized period when approximately 83 PFS events have occurred. This is an Information Fraction (IF) of 0.30 (83/278 events that are required for the primary PFS analysis).

The futility boundary will be based on the conditional probability of showing statistical significance for the primary endpoint of PFS, which is based on 278 events in approximately 556 patients. If the conditional probability that the final study result will be statistically significant, given the data observed thus far and assuming the original design effect for the remainder of the study is less than 30%, the IDMC will consider the option of declaring futility. The exact figure used for the futility boundary will be calculated by the AZ statistician or delegate and sent to the IDMC at the time of the interim analysis, based on the number of events which have occurred at that time using appropriate software such as EAST. As an example, if exactly 83 of approximately 278 final events have occurred at the time of the interim analysis, then the HR that corresponds to 30% conditional power for the interim analysis will be 1.262.

6. CHANGES OF ANALYSIS FROM PROTOCOL

The primary PRO symptoms will be analysed using a MMRM analysis of the change from baseline.

For randomized period, ORR and DoR by BICR will be repeated as exploratory endpoints.

ANCOVA analysis for the depth of response for randomized period is added.

The IDMC may propose the interv ls of the periodic safety data review according to their reviews of the randomized patients safety data.

7. REFERENCES

Berry et al 1991
Berry G, Kitchin RM, Mock PA. A comparison of two simple hazard ratio estimators based on the logrank test. Statistics in Medicine 1991;10:749-55.

Cockcroft & Gault, 1976
Cockcroft, D. W. & Gault, M. H., 1976. Prediction of creatinine clearance from serum creatinine. Nephron, 16(1), pp. 31-41.

EuroQol Group 1990
EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.

Fayers et al 2001
Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A; EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer; 2001.

Fleischer et al 2011
Fleischer F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: A quantitative approach. Stat Med. 2011;30(29):3373-86.

Gail, M. & Simon, R., 1985
Testing for qualitative interactions between tr atm nt effects and patient subsets. Biometrics, Volume 41, pp. 361-72.

Klein et al 2007
Klein JP, Logan B, Harhoff M et al. Analyzing survival curves at a fixed point in time. Stat Med. 2007; 26:4505-4519

Obosa et al 1998
Obosa D, Rodreigues G, Myles J, Zee B and Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998;16(1):139-44.

Robins 1993
Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. ASA Proceedings of Biopharmaceutical Section. 1993:24-33

Robins et al 1991
Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. Commun Stat Theory Methods. 1991;20(8):2609-31.

Selke & Siegmund 1983
Selke T, Siegmund D. Sequential analysis of the proportional hazards model. Biometrika 1983;70:315-26.

Sun and Chen 2010
Sun X, Chen C. Comparison of Finkelstein's method with the conventional approach for interval-censored data analysis. Stat Biopharm Res. 2010; 2:97-108.

Van Hout et al. 2012
Van Hout B, Janssen MF, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012; 15(5):708-15

8. DOCUMENT REVISION HISTORY

Version	Description of Update
2.0 (January 03rd 2023)	• Specified more details how to summarize subjects who switched the treatment cohort in safety run-in period (Section 2.1.1) • Defined Pharmacokinetic Analysis Set in safety run-in period (Section 2.1.1) • Added "no missing doses 7 days prior to the PK sample" in Pharmacokinetic analysis set definition (Section 2.1.2) • Modified the wording in IPD list to add clarity (Section 2.2) • Added Best overall response definition for both safety run-in period and randomized period (Section 3.2.6) • Edited Table 8 (removed BICR PFS for subjects not in cFAS, added CNS PFS by CNS status at baseline) and updated the corresponding analysis (Section 4.2.13) • Added Fatigue and Appetite loss to primary PRO measurement list (Section 3.3.1) • Updated the PRO QLQ-LC13 and QLQ-C30 cen oring window (Section 3.3.1) • Adding missing category to PRO clinically meaningf l changes (Section 3.3.1) • Added Time to definitive deterioration as sensiti ity analysis (Section 3.3.1) • Added TTD and TTDD cenoring rule tables (Table 12 and Table 13) • Deleted PRO improvement rate definition and analysis (Section 3.3.1 and Section 4.2.12) • Added dose delays for chemotherapy (Section 3.5.1 and Section 4.3.4) • Added definitions and summaries for RDI and PID (Section 3.5.1 and Section 4.3.4) • Added "date of first subsequent a ti-cancer therapy" as one of the cutting date in safety follow-up period definition (Section 3.5.1) • Added one more grouped term (Hematological Toxicity Grouped Terms (Narrow)) to AESI (Section 3.5.3) • Additional conditi n is add d when considering the time of the primary analysis for the PFS, i.e. The pri ary analysis of PFS based on Investigator assessment (according to RECIST 1.1) will oc ur when approximately 278 PFS events (approximately 50% maturity) and a least 16 months of follow-up after LSI has occurred in the 556 randomized patients.[update accordingly for the Amendment of CSP v2] • For all Time to event analysis, change the breslow method to Efron method for handling ties (Section 4.2) • Added K-M plots and estimates for OS in safety run-in period (Section 4.2.3.1 and 4.2.4.1) • Added quantitative interaction sensitivity analysis for primary PFS (Section 4.2.3.2) • Added K-M curves and K-M estimates for the randomization phase DoR (Section 4.2.6) • Added wording to clarify how to define the CNS status at baseline (Section 4.2.9) • Added log-rank test for TTD of each PRO subscales (Section 4.2.12) • Modified the analysis visit mapping window for PRO QLQ-LC13 and QLQ-C30 (Section 4.2.12) • Added safety assessment analysis visit mapping window for Japan site (Section 4.3.1.1) • Added more AESI summaries (Section 4.3.1.3) • Deleted the management of AEs of diarrhea and rash (Section 4.3.1.3) • Further details on assessments of Impact of COVID-19 (Section 4.3.6)

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: D5169C00001-sap-ed-2.0		
Document Title:	D5169C00001 Statistical Analysis Plan Edition 2.0	
Document ID:	Doc ID-005064722	
Version Label:	1.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
09-Jan-2023 10:07 UTC	PPD	Content Approval
09-Jan-2023 17:01 UTC	PPD	Author Approval
09-Jan-2023 16:57 UTC	PPD	Qualified Person Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.