Statistical Analysis Plan

Study Code	D5160C00048
Edition Number	3.0
Date	13-Dec-2023

A Phase III, randomized, double-blind, placebo-controlled, multicenter, international study of osimertinib as maintenance therapy in subjects with locally advanced, unresectable EGFR mutation-positive Non-Small Cell Lung Cancer (Stage III) whose disease has not progressed following definitive platinum-based chemoradiation therapy (LAURA)

TABLE OF CONTENTS

TITLE PAGE....1
TABLE OF CONTENTS....2
LIST OF ABBREVIATIONS....5
AMENDMENT HISTORY....9
1 STUDY DETAILS....36
1.1 Study objectives....36
1.1.1 Primary Objective....36
1.1.2 Secondary Objectives....36
1.1.3 Exploratory Objectives....37
1.2 Study design....39
1.3 Number of subjects....42
2 ANALYSIS SETS....43
2.1 Definition of analysis sets....43
2.1.1 Full analysis set....43
2.1.2 Safety analysis set....43
2.1.3 Pharmacokinetic analysis set....43
2.1.4 Evaluable for response analysis set (subset of Full Analysis Set)....44
2.2 Protocol deviations....44
3 PRIMARY, SECONDARY, AND EXPLORATORY VARIABLES....46
3.1 Derivation of RECIST visit response....46
3.1.1 Blinded Independent Central Review (BICR) Assessment using RECIST v1.1....47
3.1.1.1 CNS BICR....48
3.1.2 Site Investigator Assessment Using RECIST v1.1....49
3.1.2.1 Target lesions (TLs)....49
3.1.2.2 Non-target lesions (NTLs) and New lesions....53
3.1.2.3 New lesions....54
3.1.3 Overall visit response....55
3.2 Efficacy outcome variables....55
3.2.1 Primary outcome: progression-free survival (PFS)....55
3.2.2 Secondary outcomes....57
3.2.2.1 CNS progression-free survival (CNS PFS)....57
3.2.2.2 Overall survival (OS)....57
3.2.2.3 Objective response rate (ORR)....57
3.2.2.4 Duration of response (DoR)....58
3.2.2.5 Disease control rate (DCR)....58
3.2.2.6 Best Objective Response (BoR)....58
3.2.2.7 Depth of response....59
3.2.2.8 Time to death or distant metastases (TTDM)....59
3.2.2.9 Time to study treatment discontinuation or death (TTD)....60
3.2.2.10 Time from randomization to second progression-free survival or death (PFS2)....60
3.2.2.11 Time to first subsequent therapy or death (TFST)....60
3.2.2.12 Time to second subsequent therapy or death (TSST)....61

3.3 Patient-reported outcomes (PROs) 61
3.3.1 EORTC QLQ-C30 and EORTC QLQ-LC13 61
3.3.2 Patients Global Impression of Severity (PGIS) 65
3.3.3 Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Event (PRO-CTCAE) 65
3.3.4 EuroQoL 5-Dimension 5-Levels (EQ-5D-5L) 65
3.3.5 Health resource use module 66
3.4 Safety assessments 66
3.4.1 General considerations for safety assessments 66
3.4.1.1 Safety Follow-up 66
3.4.1.2 Handling missing data 66
3.4.2 Exposure and dose interruptions 68
3.4.3 Laboratory assessments 68
3.4.4 Physical examinations 68
3.4.5 Vital signs 68
3.4.6 Electrocardiograms (ECGs) 69
3.4.7 Left ventricular ejection fraction (LVEF) 69
3.4.8 WHO performance status 69
3.4.9 Adverse events (AEs) 69
3.5 Pharmacokinetics 70
3.6 Other exploratory variables 70
4 ANALYSIS METHODS 71
4.1 General principles 72
4.1.1 Visit windows 73
4.2 Analysis methods 74
4.2.1 Demographic and other baseline characteristics 74
4.2.2 Protocol Deviations 76
4.2.3 Concomitant and other treatments 76
4.2.4 Exposure 77
4.2.5 Subject disposition and data sets analyzed 77
4.2.6 COVID-19 Impact 78
4.2.7 Primary outcome: PFS 78
4.2.7.1 Primary analysis of PFS 78
4.2.7.2 Sensitivity analyses of PFS 79
4.2.7.3 Subgroup analyses of PFS 81
4.2.8 Secondary outcomes 83
4.2.8.1 CNS progression-free survival (CNS PFS) 83
4.2.8.2 Overall survival (OS) 85
4.2.8.3 Sensitivity Analysis – Accounting for COVID-19 deaths 85
4.2.8.4 Objective response rate (ORR) 85
4.2.8.5 Duration of response (DoR) 86
4.2.8.6 Disease control rate (DCR) 86
4.2.8.7 Depth of response 86
4.2.8.8 Time to death or distant metastases and time to study treatment discontinuation (TTDM and TTD). 87
4.2.8.9 Post progression outcomes (PFS2, TFST and TSST) 88
4.2.9 Patient reported outcomes 88
4.2.9.1 EORTC QLQ-C30 and EORTC QLQ-LC13 88

4.2.9.2 Subject Reported Outcome version of the Common Terminology Criteria for Adverse Event System (PRO-CTCAE) 90
4.2.9.3 Patients Global Impression of Severity (PGIS) 90
4.2.9.4 EQ-5D-5L health state utility index 90
4.2.9.5 Healthcare resource use module 90
4.2.10 Safety 90
4.2.10.1 Laboratory assessments 91
4.2.10.2 Physical examinations 92
4.2.10.3 Vital signs 92
4.2.10.4 Electrocardiograms (ECGs) 92
4.2.10.5 Left ventricular ejection fraction (LVEF) 92
4.2.10.6 WHO performance status 92
4.2.10.7 Adverse events (AEs) 93
4.2.11 Pharmacokinetic data (PK) 96
4.2.12 Other exploratory outcomes 97
4.2.12.1 Relationship between PK and other variables 97
4.2.12.2 Comparison of baseline tumor EGFR mutation status 97
5 INTERIM ANALYSES 99
5.1 Independent Data Monitoring Committee (IDMC) 99
5.2 Interim analysis for OS 99
6 REFERENCES 99

LIST OF TABLES

Table 1 Primary Objective 36
Table 2 Secondary Objectives 36
Table 3 Exploratory Objectives 37
Table 4 Summary of outcome variables and analysis set 44
Table 5 TL visit responses (RECIST v1.1) 49
Table 6 NTL visit response (RECIST v1.1) 53
Table 7 Overall visit responses 55
Table 8 Visit responses for symptoms and HRQoL 63
Table 9 2x2 table format for reporting results comparing a new test outcome to the reference standard outcome 98

LIST OF FIGURES

Figure 1 Overall study design 42

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	Alanine aminotransferase
AST	Aspartate aminotransferase
AUC	Area Under the Curve
BP	Blood Pressure
BICR	Blinded Independent Central Review
BoR	Best objective Response
CCRT	Concurrent Chemoradiation
CDx	Companion Diagnostic
CI	Confidence Interval
CLIA	Clinical Laboratory Improvement Amendments
CNS	Central Nervous System
COVID-19	Coronavirus 2019 disease
CR	Complete Response
CRO	Clinical Research Organization
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT scan	Computerized Tomography scan
CTCAE	Common Terminology Criteria for Adverse Events
ctDNA	circulating tumor DNA
DBL	Database Lock
DCO	Data Cut-Off
DCR	Disease Control Rate
DNA	Deoxyribonucleic Acid
DoR	Duration of Response
ECG	Electrocardiogram
EDR	Early Discrepancy Rate
EFR	Evaluable for Response Analysis Set
EGFR	Epidermal Growth Factor Receptor
EORTC	European Organization for Research and Treatment of Cancer

Abbreviation or special term	Explanation
EQ-5D-5L	EuroQoL 5-Dimension 5-Level health state utility index
Ex19Del	Exon 19 Deletion
FAS	Full Analysis Set
FDA	Food and Drug Administration
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hazard Ratio
HRQL, HRQoL	Health-Related Quality of Life
HRU	Health Resource Use
HL	Hy's Law
ICF	Informed Consent Form
ICH	International Conference on Harmonization
ICR	Independent Review Charter
ITT	Intent-To-Treat
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
KM	Kaplan-Meier
L858R	An amino acid substitution at position 858 in EGFR, from a leucine (L) to a arginine (R)
LD	Longest Diameter
LDR	Late Discrepancy Rate
LRCI	Likelihood Ratio-Based Confidence Intervals (SAS^{Ò} syntax)
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
MET	tyrosine-protein kinase Met
MMRM	Mixed effect Models for Repeated Measures
MRI	Magnetic Resonance Imaging scan
MSSO	(ICH MedDRA) Maintenance and Support Services Organization
MTP	Multiple Testing Procedure
MUGA	Multi Gated Acquisition Scan
NCI	National Cancer Institute
NE	Not Evaluable
NED	No Evidence of Disease
NPA	Negative Percent agreement

Abbreviation or special term	Explanation
NPV	Negative Predicted Value
NSCLC	Non-Small Cell Lung Cancer
NTL	Non-Target Lesion
OPA	Overall Percent Agreement
ORR	Objective Response Rate
OS	Overall Survival
PFS	Progression-Free Survival
PFS2	Time from randomization to Second Progression on subsequent treatment
PGIS	Patients Global Impression of Severity
PK	Pharmacokinetic(s)
PPA	Positive Percent Agreement
PPV	Predictive Value of a Positive result
PR	Partial Response
PRO	Patient Reported Outcome
PRO-CTCAE	Patient Reported Outcome - Common Terminology Criteria for Adverse Event
QD	Quaque Die (Once Daily)
QLQ-C30	30-Item Core Quality of Life Questionnaire
QLQ-LC13	13-Item Lung Cancer Quality of Life Questionnaire
QoL	Quality of Life
QTcF	Corrected QT Interval by Fridericia
RECIST	Response Evaluation Criteria in Solid Tumors
REML	Residual Maximum Likelihood
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SCRT	Sequential Chemoradiation
SD	Stable Disease
TL	Target Lesion
TFST	Time to First Subsequent Therapy
TSST	Time to Second Subsequent Therapy

Abbreviation or special term	Explanation
TTD	Time to Treatment Discontinuation or Death
TTDM	Time to Death or Distant Metastases
USA	United States of America
VCV	Veeva Clinical Vault
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

AMENDMENT HISTORY

Category*:	Date	Description of change	In line with the CSP?	Rationale
Change refers to				
Version 3.0				
Multiple testing procedure	13-Dec-2023	Section 4 – "Multiple testing strategy" updated to describe the updated hierarchical order of the endpoints in the MTP: PFS -> OS -> CNS PFS Throughout – wording of 'CNS PFS and OS' changed to 'OS and CNS PFS' to reflect updated MTP order.	Y (v5.0)	To align with the MTP update that was made to CSP v5.0.
Primary or secondary endpoints	13-Dec-2023	Section 3.2.1, 3.2.2, 4.2.7.2.3, 4.2.7.2.4 last evaluable RECIST assessment changed to last non-missing RECIST assessment.	N	As subjects are randomized after they have had CRT, there is a higher chance that subjects will still have radiation effects that obscure the investigator/BICR from being able to properly evaluate the RECIST assessments, leading to NE (not evaluable) assessments.
Primary or secondary endpoints	13-Dec-2023	Section 3.2.1 2 missed visit rule corrected.	N	To align with AZ standards
Primary or secondary endpoints	13-Dec-2023	Section 3.2.2.1 Wording updated to align with Section 3.1.1 updates.	N	Clarification.
Primary or secondary endpoints	13-Dec-2023	Section 3.2.2.10 Clarification added that PFS2 uses the investigator assessment of first PFS, even though the primary endpoint is by BICR.	N	To align with the TA SAP

Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.1 <10 subjects changed to events. Estimates and 95% CI at 6monthly intervals added. Prematurely censored subjects text removed.	N	To align with TA SAP and further summarize the primary endpoint.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.2.1 Added sensitivity analysis for stratification according to the eCRF.	N	To align with TA SAP in the situation where a study has mis-stratifications.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.2.2 KM plot added for ascertainment bias.	N	To support interpretation.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.2.5 and 4.2.8.3 added sensitivity analysis for COVID-19 for PFS and OS.	N	As per TA SAP.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.3.1 Added Chinese vs non-Chinese subgroup for subgroup analysis.	Y (v4.0)	To align with CSP.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.3.1 Modified cox proportional hazards model.	N	As per TA SAP.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.3.1 Added condition that subgroups will not be formally analyzed when a subgroup category has less than 20 events across both treatment groups.	N	As per TA SAP.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.1 A listing of all tumor assessment based on investigator and central review added. Added estimates and 95% CI for PFS rates at 6 monthly intervals.	N	To support interpretation
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.7.1 and 4.2.8.1 added summary of disease characteristics and summary of disease characteristics at CNS progression.	N	To support interpretation.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.1 Added KM plot for CNS PFS. Added CNS progression summary table. Added treatment status at CNS progression summary table.	N	To support interpretation of the CNS PFS endpoint

Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.2 Added KM plot. Prematurely censored summary removed.	N	To support interpretation of the OS endpoint
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.4 Separation between confirmed vs unconfirmed BOR and clarification added for analysis set.	N	Clarification.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.5 Analysis set and repeated output for confirmed BICR added. Clarification on BICR data added.	N	To support interpretation of response data
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.7 Details added for waterfall plot figure. Text referencing scaled measurements removed. Not applicable for BICR data.	N	Clarification.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.8, 4.2.8.9 Clarified that same method as primary analysis of PFS is applied to obtain the HR and CI.	N	Clarification.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.8.8 Conditional probability removed.	N	Cumulative incidence sufficiently summarizes Competing Risk analysis.
Statistical analysis method for the primary or secondary endpoints	13-Dec-2023	Section 4.2.9.1 MMRM limited to data from baseline up to 10 months or date of PD.	N	To perform the analysis on an appropriate timeframe of data to provide more meaningful outcome of the analysis.
Derivation of primary or secondary endpoints	13-Dec-2023	Section 3.2.2.6 Clarifications added on the derivation of BoR.	N	Clarification.
Derivation of primary or secondary endpoints	13-Dec-2023	Section 4.2.7.2.3 Clarified censoring information for evaluation-time bias.	N	To align with the TA SAP.
Derivation of primary or secondary endpoints	13-Dec-2023	Section 4.2.8.1 and 4.2.8.8 Clarification added for subjects who die in the absence of progression.	N	Clarification.

Derivation of primary or secondary endpoints	13-Dec-2023	Section 3.4.1.1 and Section 4.2.10.7 Added text to clarify that palliative radiotherapy is not counted as a subsequent anti-cancer therapy when calculating safety follow-up period.	N	To align with the TA SAP
Data presentation	13-Dec-2023	Section 2 Updated to clarify Table 4 and add mis-stratifications tables.	N	Clarification.
Data presentation	13-Dec-2023	Section 2.15 Updated to include only the PFS and ORR sensitivity analysis for the Evaluable for response analysis set .	Y (v4.0)	To align with CSP.
Data presentation	13-Dec-2023	Section 3.3.1 Added text to clarify that all symptoms, functioning subscales, and the GHS/QoL subscale from the QLQ-C30 should be included in the time to symptom deterioration analysis. Updated time to symptom/function/QoL deterioration definition to match with protocol.	N	Clarification.
Data presentation	13-Dec-2023	Section 4.1 Baseline definition for PRO endpoints modified	N	To align with TA SAP definition
Data presentations	13-Dec-2023	Section 4.1.1 Updated minimum of number of observations required for by-visit tables.	N	To focus outputs on visits where meaningful data interpretation can be made
Data presentation	13-Dec-2023	Section 4.2.1 Added mean lung radiotherapy dose and v5, v20, Cardiac v30, v45 and v50 at baseline, time from unresectable Stage III diagnosis to randomization, baseline TL size by investigator, response to prior chemoradiation (SD vs PR vs CR), disease related medical history, surgical history.	N	To further summarize the disease characteristics of the study population
Data presentation	13-Dec-2023	Section 4.2.1 and 4.2.3 added COVID-19 summaries and listings for IPD, baseline characteristics, medical histories and ongoing conditions.	N	To further support the summaries of the study population

Data presentation	13-Dec-2023	Section 4.2.2 added.	N	To ensure the planned summaries of Protocol Deviations are clear
Data presentation	13-Dec-2023	Section 4.2.3 Clarified that prior medication includes radiotherapy.	N	Clarifications.
Data presentation	13-Dec-2023	Section 4.2.4 Added duration of exposure summary for those receiving open-label Osimertinib.	N	To further summarize open-label Osimertinib data
Data presentation	13-Dec-2023	Section 4.2.6, 4.2.10.7 Added details for summarizing COVID-19 related data.	N	To align with TA SAP
Data presentation	13-Dec-2023	Section 4.2.8.9 Added KM plot for PFS2, TFST and TSST.	N	To support interpretation of the endpoints
Data presentation	13-Dec-2023	Section 4.2.9.1 Absolute score and change from baseline added to ‘response by visit’ output. 'All visit' was replaced with 'key visits'	N	To further summarize the PRO data
Data presentation	13-Dec-2023	Section 4.2.9.2 Branching logic added.	N	Branching logic to be applied when summarizing the data
Data presentation	13-Dec-2023	Section 4.2.9.3 and 4.2.9.4 Key visits for inclusion in summaries were listed.	Y (v4.0)	To avoid producing summaries at visits with few observations
Data presentation	13-Dec-2023	Section 4.2.10 Baseline definition for safety data moved from Section 4.2.10.1 to 4.2.10.	N	Clarification.
Data presentation	13-Dec-2023	Section 4.2.10.1 Added ALP elevation for Hy’s Law output.	N	To further assess lab data
Data presentation	13-Dec-2023	Section 4.2.10.3/4 Added box plot for absolute and change from baseline values for vital signs and ECG.	N	To support interpretation of safety data
Data presentation	13-Dec-2023	Section 4.2.10.4 added text for ECG listing.	N	To support interpretation of ECG data

Data presentation	13-Dec-2023	Section 4.2.10.5 Added text for LVEF listing.	N	To support interpretation of LVEF data
Data presentation	13-Dec-2023	Section 4.2.10.6 Added text for WHO performance status listing.	N	To support interpretation of WHO performance data
Data presentation	13-Dec-2023	Section 4.2.10.7 Added SAE summary for subjects receiving open-label Osimertinib.	N	To support interpretation of open-label Osimertinib data
Data presentation	13-Dec-2023	Section 4.2.10.7 Corrected 90 days after last dose to 28 and death outputs to be presented based on FAS.	N	To correct errors.
Data presentation	13-Dec-2023	Section 4.2.10.7 Added cardiac events listing for CTCAE grade 2 or higher. Added time to onset from last dose of radiotherapy summaries. Added AESI summary table for subjects receiving open-label Osimertinib. Added text for AESI prevalence plots.	N	To further support the analysis of safety data
Data presentation	13-Dec-2023	Section 4.2.11 A listing of individual plasma/serum concentration for each treatment added.	N	To support interpretation of PK data
Data presentation	13-Dec-2023	Section 4.2.12.2 Clarified comparison of baseline tumor EGFR mutation status will only be performed if sufficient EGFRm data available.	N	Clarification.
Other	13-Dec-2023	Tumour changed to Tumor (change from UK to US English).	N	Correction.
Other	13-Dec-2023	Section 1.2, figure 1 footnotes and section 1.3.	Y (v5.0)	To align with CSP.
Other	13-Dec-2023	Section 1.2 Updated text for how long AZ will continue to supply open-label Osimertinib after OS	Y (v3.0)	To align with CSP.
Other	13-Dec-2023	Section 2.1.1 and 2.1.2 update to FAS and SAF definition.	Y (v4.0)	To align with CSP.
Other	13-Dec-2023	Section 3.1.1.1 was added to clarify CNS BICR assessment is separate from whole body BICR	N	Clarification.

Other	13-Dec-2023	Section 3.1.1, 3.2.2.8 Updated definition of 'distant' metastases from the definition in the CSP	N	Clarification for how a distant metastases is defined for the study
Other	13-Dec-2023	Section 3.3 Added text to clarify post-PD PRO follow-up and text to clarify when PRO analysis is to take place.	Y (v4.0)	For clarification and to align with CSP
Other	13-Dec-2023	Section 3.3.1 Added subjects evaluable for time to deterioration analysis as subjects with baseline >=10. Updated definition for deterioration.	Y (v4.0)	To align with CSP.
Other	13-Dec-2023	Section 3.3.1 Updated text from HRQL to HRQoL.	N	For consistency of naming throughout
Other	13-Dec-2023	Section 3.3.1 Clarified which items are to be included for time to symptom deterioration.	N	Clarification.
Other	13-Dec-2023	Section 3.3.1 Updated time to symptom/function/QoL deterioration definition to included confirmation at the next consecutive visit.	N	To align with AZ Oncology PRO Guidance.
Other	13-Dec-2023	Section 3.4.1.2 Added text on missing date imputation for subsequent anti-cancer therapy.	Y (v4.0)	To align with CSP.
Other	13-Dec-2023	Section 3.4.6 Mean of available ECG results added.	N	Clarification.
Other	13-Dec-2023	Section 3.4.9 Updated collection of AE and SAE information.	Y (v4.0)	To align with CSP.
Other	13-Dec-2023	Section 3.4.9 Added text about COVID-19 related AE. Cardiac effects added as programme-wide AESI.	N	Clarification.
Other	13-Dec-2023	Section 4.2.1 Added in summaries and listings for medical and surgical history.	N	To support interpretation of data.
Other	13-Dec-2023	Section 4.2.7.1 Clarified which symptom/functional items MMRM will be performed on.	N	Clarification.

Other	13-Dec-2023	Section 4.2.9.1 Updated text from HRQL to GHS/QOL for the response by visit sub section.	NA	Clarification.
Other	13-Dec-2023	Section 5.2 Reference to Section 4 added.	N	Clarification.
Other	13-Dec-2023	Amendment History Table: The date of changes made to Version 2.0 has been updated to be in line with the date of SAP v2.0	N	Correction of date error.
Version 2.0				
Primary or secondary endpoints	27-May-2020	Section 3.2.6, "Best Objective Response" added to DCR section	N	Further assessment of efficacy
Primary or secondary endpoints	27-May-2020	Section 2.1.4 – Removed DoR from list of endpoints that will be repeated in the FAS set. Table 4 - Removed DoR from list of endpoints that will be repeated in the FAS set and included in list of endpoints that will be evaluated in the evaluable for response analysis set. Section 4.2.6.4 – Add text to clarify that DoR only assessed in EFR set.	N/A	DoR can only be assessed in subjects who respond.
Primary or secondary endpoints	27-May-2020	Section 4.2.6.1 – Added competing risks section for CNS progression.	N	To account for competing events in evaluation of CNS progression.
Primary or secondary endpoints	27-May-2020	Section 4.2.6.7 – Added competing risks for TTDM	N	To account for competing events in evaluation of TTDM.
Primary or secondary endpoints	27-May-2020	Section 3.2. Removed text: "CR or PR must be confirmed"	Y (v1.0)	Not a condition that response requires confirmation. Analysis is repeated for confirmed response.
Primary or secondary endpoints	27-May-2020	Section 1.1.2 (Table 2). "EGFRm+ Ex19del or L858R detectable in plasma-derived ctDNA" to	Y (v2.0)	To align with CSP.

		"EGFR Ex19del or L858R mutation detectable in plasma-derived ctDNA"		
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.5.1. Added detail on how to handle non-proportional hazards.	N/A	As per the TA SAP.
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.5.1. Added detail on which baseline prognostic factors will be presented for subjects prematurely censored.	N/A	Clarification.
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.5.1.2. Corrected text to state that 'A negative differential discordance for the EDR and/or positive differential discordance for the LDR are suggestive of a bias in the BICR assessment favoring the experimental arm' rather than 'Investigator'.	N/A	Correction.
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.6.3. Updated text to confirm that in the main analysis of ORR, both confirmed and unconfirmed responses will be used.	N/A	Clarification
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.6.4. Analysis of DoR updated to align with the TA SAP at client request.,	N/A	To align with TA SAP.
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.7.1. Response by visit summaries added to enable interpretation of formal MMRM analysis.	N/A	At client request.
Statistical analysis method for the primary or secondary endpoints	27-May-2020	Section 4.2.5 and 4.2.6. Removed any reference to the 'Breslow approach for handling ties' in relation to the log-rank test.	N/A	To align with TA SAP.
Derivation of primary or secondary endpoints	27-May-2020	Section 3.1.2.1 – Fourth and fifth paragraphs added; Under "Missing TL data", third bullet point "(the previous minimum sum of diameters)" was added; Under "Lesions that merge", updated from 0cm to 0mm	Y (v1.0)	To describe how Overall Visit Response will be determined for subjects without measurable disease at baseline

Derivation of primary or secondary endpoints	27-May-2020	Table 7 updated to include criteria for when TLs are NA NTLs are CR, Non-CR/Non-PD, NE or NA and there are no new lesions	Y (v1.0)	Required for determining Overall Visit Response
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.1 PFS – Third paragraph updated to be 16+1 week	Y (v1.0)	To account for late assessments
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.1 PFS – First bullet point under fourth paragraph updated to read "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 selected single reviewer described in 3.1.1"	Y (v1.0)	RECIST scans/assessments contributing to response may occur on separate dates
Derivation of primary or secondary endpoints	27-May-2020	Sections 3.2.2.10 (TFST) and 3.2.2.11 (TSST) updated in line with TA SAP	N/A	In line with TA SAP
Derivation of primary of secondary endpoints	27-May-2020	Section 4.2.8.2 updated to specify the calculations for PPA, NPA, OPA, PPV, NPV and 95% CI for each	N/A	To provide further detail
Derivation of primary of secondary endpoints	27-May-2020	Section 3.1 – Added text: "Progression includes disease recurrence for subjects with no evidence of disease at baseline."	N	Text added for clarification.
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.2.5 – Added underlined text: "BoR will be determined programmatically based on RECIST from the overall visit response using all BICR data up until the first progression event or subsequent therapy, whichever is the earliest."	Y (v1.0)	Text added for clarification.
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.2.6 – Added text from TA SAP detailing how to handle calculation of best percentage change in the absence of missing data.	N/A	In line with TA SAP.

Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.2.10 – Removed sentence “This will exclude anti-cancer surgery as a subsequent therapy”	Y (v1.0)	Surgery very rare in the subject population. Updated text consistent with similar studies (FLAURA).
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.2.2 – Updated text to include guidance on the imputation of partial death dates for OS.	N/A	In line with the TA SAP.
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1 – Updated to include a month is defined as 30.4375 days	N/A	Clarification where month is used in any derivations.
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1 – Added text on the calculation of change from baseline for the safety endpoints.	N/A	To provide further detail (in line with TA SAP)
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1 – Added text "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."	Y (v1.0)	To provide further detail on how to handle assessments where actual or nominal time is not recorded, in regard to baseline.
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1 – Added text to clarify that baseline is the last value obtained prior to randomization for efficacy endpoints.	Y (v1.0)	Clarification.
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1 – Added text to clarify timing of baseline assessments for PROs	Y (v1.0)	Clarification.
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1.1 – Clarification of final bullet to indicate that **best** value will be used as baseline where there are multiple eligible visits for non-numeric laboratory, rather than average. Average to be used for numeric parameters.	N/A	Clarification.
Derivation of primary or secondary endpoints	27-May-2020	Section 4.1.1. – Updated to add text “If there are two equally eligible assessments at a visit, then the worst value will be taken as the post-baseline value”	N/A	To clarify what to do in the case where 2 post-baseline values are equally eligible. Taking the worst value is the most conservative approach.

Derivation of primary or secondary endpoints	27-May-2020	Section 4.2.8.7 – Added text: "If a subject has multiple fatal AEs, the earliest start date is used to determine the category they fall into, i.e. onset < 90 days after last dose of study medication or onset > 90 days after last dose of study medication."	N/A	Clarification of which start date to use in case where subject has multiple AEs recorded as fatal.
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.1. Added text to clarify that a response of 'NE' is not considered a missed visit when considering the 2 or more missed visit censoring rule for PFS.	N/A	As per the TA Oncology SAP guidance v4
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.1. Corrected spelling in final two bullet points from "investigational assessments" to "investigator assessments".	N/A	Correction required
Derivation of primary or secondary endpoints	27-May-2020	Section 3.4.2. Added a new section on "Exposure and dose interruptions" and added detail on the derivation of exposure duration. Remaining sections 3.4.x shifted down one place.	N/A	As per the TA Oncology SAP guidance v4
Derivation of primary or secondary endpoints	27-May-2020	Section 3.2.2.11. Update to definition of TSST replacing 'first subsequent treatment' with 'randomized treatment' which is more technically correct.	N/A	Correction.
Multiple testing procedure	27-May-2020	Section 4 – "Multiple testing strategy" updated to include information on significance level for interim and final analysis	Y (v1.0)	To provide further detail
Data presentations	27-May-2020	Section 4.2.6.7 – Updated to include summary of AEs leading to dose reduction, dose interruption, discontinuation of study medication and possibly related to IP	N/A	Client request

Data presentations	27-May-2020	Section 4.2.6.6 – Added text: "Values of depth of response or best percentage change from baseline that have been calculated using scaled measurements will be identified and flagged with an appropriate symbol, clearly indicated on the plot or listing."	N	Client request to highlight values based on scaled measurements.
Data presentations	27-May-2020	Section 4.1 – Added bullet that missing coded terms should be listed and summarized as 'Uncoded'	N/A	As per TA SAP guidance.
Data presentations	27-May-2020	Added Concomitant and other treatment section (4.2.2) to describe these summaries.	N/A	As per TA SAP.
Data presentations	27-May-2020	Section 4.2.5.2.2 – Removed "This analysis will be repeated for subjects with EGFR Ex19del or L858R mutation and also for subjects with EGFRm+ Ex19del or L858R detectable in plasma-derived ctDNA."	N/A	Not required.
Data presentations	27-May-2020	Section 4.2.5.3 – Only stratification factors considered for subgroup analysis. All other subgroups removed. Also documented in Section 6 – Changes of Analysis from Protocol	N	Given the small size of the study, the number of subgroups to be explored has been reduced.
Data presentations	27-May-2020	Section 4.2.8.7 – Added summary of AEs leading to dose modification.	N/A	Required for IDMC charter.
Data presentations	27-May-2020	Section 4.2.5.3 – Central EGFR and tissue mutation type added back into list of subgroup analysis.	N	Secondary objective
Data presentations	27-May-2020	Section 4.1.1 – Updated minimum number of observations required for by-visit plots and removed restrictions on tables.	N/A	All data to be presented in tables, but figures to have minimum of 5 observations to prevent very large plots with meaningless data.

Data presentations	27-May-2020	Section 4.2.8.1. Added summaries for Hy's Law.	Y (v2.0)	As per TA Oncology SAP guidance v4.0.
Data presentations	27-May-2020	Section 3.4.8. Amended text to state that in addition to SAEs, Adverse Events of Special Interest (AESIs) will also be recorded to post the final OS analysis if applicable.	Y (v2.0)	To align with CSP.
Data presentations	27-May-2020	Section 4.2.1. Added 'Important protocol deviations' as a bullet under categorical variables.	N	Client request.
Data presentations	27-May-2020	Section 4.2.8.1. Updated urinalysis categories.	N	Client request.
Data presentations	27-May-2020	Section 4.2.1. Additional categorical variables added	Y (2.0)	In line with eCRFs and TA SAP.
Data presentations	27-May-2020	Section 4.2.4. Screening failures added to subject disposition summaries.	Y (2.0)	In line with eCRFs and TA SAP.
Other	27-May-2020	Replaced all occurrences of the word "patient(s)" with "subject(s)"	N	To be consistent with ICH E6 and TFLs
Other	27-May-2020	Abbreviations added to the list	N/A	In line with use within this SAP
Other	27-May-2020	Table footers updated/added and layout updated	N/A	For consistency throughout SAP
Other	27-May-2020	The wording "analysis population(s)" was replaced with "analysis set(s) throughout	Y (v1.0)	For consistency with CSP
Other	27-May-2020	Abbreviation for safety analysis set has been updated from SAS to SAF	N	Due to SAS being the statistical software used for analysis.

Other	27-May-2020	Section 1.2 – First sentence added to reference CSP version 1.0; The following sentence was added to the third paragraph "For these subjects investigators will be asked to provide an archival formalin-fixed and paraffin embedded (FFPE) tumor tissue sample, where available, for retrospective central analysis of EGFR mutation status in Part II screening for comparison with the local test. However subject enrollment will be based upon the local test."	Y (v1.0)	To provide further details
Other	27-May-2020	Section 1.3 – Third paragraph updated to clarify that CR, PR and SD are evaluated by the investigator according to RECIST 1.1; Fifth paragraph updated to include PFS assumption for placebo “assuming a median of 8 months PFS for placebo.”	Y (v1.0)	To provide further details.
Other	27-May-2020	Section 3.1 – Fourth paragraph updated to read “until objective radiological disease progression as defined by RECIST 1.1 and as confirmed by BICR.”; Fifth paragraph updated to read “study treatment and tumor assessments should be continued”	Y (v1.0)	To provide further details.
Other	27-May-2020	Section 3.1.1– First paragraph updated to incorporate “Information about prior radiotherapy from the CRF will also be provided to the BICR to delineate the anatomical region of previous definitive radiotherapy, to confirm that there are no lesions at baseline outside the prior radiotherapy region, to allow the selection of appropriate target lesions, if any at baseline, and to designate new lesions at follow-up that are outside the region of prior radiotherapy as ‘distant’ metastases for derivation of TTDM.”	Y (v1.0)	To provide further details.

Other	27-May-2020	Section 3.2.2.3 – Updated to incorporate definition of confirmed CR/PR	Y (v1.0)	To provide further detail.
Other	27-May-2020	Section 3.3.1 – Under "Definition of compliance and evaluability rates", sentence added to first paragraph "If compliance drops below 85%, a check-in call from the site to ask the subject if he/she has any difficulties is highly recommended."	Y (v1.0)	To provide further details.
Other	27-May-2020	Section 3.4.1.1 – "Visit windows" moved to section 4.1.1; New section 3.4.1.1, "Safety Follow-up" states "Total Safety Follow-up = min((last dose date + 28 days), date of withdrawal of consent, date of death, date of DCO, date of first dose of subsequent anti-cancer therapy) – first dose date +1."	N/A	To provide further details.
Other	27-May-2020	Section 3.4.1.2 – information added on imputation of missing dates as per TA SAP	N/A	To provide further details
Other	27-May-2020	Section 3.6 – the following was added to the first paragraph "However, the results from this research will be reported either in the CSR itself or as an addendum, or separately in a scientific report or publication may be reported separately from the CSR."	N/A	To provide further details of reporting of exploratory analysis.
Other	27-May-2020	Section 4.2.4.3 – bullet point 6 updated to include "vs unknown"	N/A	To account for subjects with unknown mutation status at screening.
Other	27-May-2020	List of abbreviations: Added CSP, ICH, EFR, removed PhUSE	N/A	As used in SAP
Other	27-May-2020	Minor updates to typos, grammar and sentence structure throughout.	N/A	To correct errors and improve reading of SAP

Other	27-May-2020	Section 1.2 – Updated to add underlined text "The final analysis of OS will be conducted at approximately 60% maturity (based on 200 subjects being randomized) when approximately 120 death events (across both treatment arms) have occurred."	Y (v1.0)	To add further detail.
Other	27-May-2020	Section 1.3 – Added underlined text: "Approximately 200 subjects will be randomized in a 2:1 ratio (osimertinib to placebo). If fewer than 40 of these subjects have been recruited in mainland China at the time of the global analysis, recruitment will continue in mainland China until approximately 40 subjects have been randomized	Y (v1.0)	To clarify that the global analysis can include subjects recruited in mainland China if they are within first 200 recruited globally.
Other	27-May-2020	Section 1.3 – Added underlined text: In order to reflect global clinical practice, recruitment will be monitored on an ongoing basis and will be managed to ensure that the majority (approximately ≥60%) of subjects entering the study have received prior CCRT	N	Wording updated so that subjects do not need to be excluded in the case that % is very close to 60% but not above.
Other	27-May-2020	Section 1.3 – Added underlined text: The primary analysis of PFS and analysis of PFS in the China cohort will occur when approximately 120 and approximately 24 PFS BICR events have been observed, respectively.	Y (v1.0)	To make it clear that 24 PFS events is an approximate number.
Other	27-May-2020	Section 1.3 – Added underlined text: The final analysis of OS will be conducted at approximately 60% maturity at approximately 120 and approximately 24 death events respectively for the final analysis and the final analysis in the China cohort.	Y (v1.0)	To make it clear that numbers are approximate.

Other	27-May-2020	Section 2.1.2 – Updated to redefine SAF as subjects who had at least one dose of study treatment. Removed text “and for whom post dose data are available”	N	In line with TA SAP
Other	27-May-2020	Section 2.1.3 – Updated definition of PK analysis set to be a subset of SAF rather than FAS.	N	Subjects must have at least one dose to have quantifiable concentration.
Other	27-May-2020	Section 2.1.4 – Updated to remove text “who have received at least one dose of treatment”. Evaluable for response set should be a subset of FAS.	N	Given EFR is used for efficacy, population should be based on FAS.
Other	27-May-2020	Section 3.1 – Added underlined text to describe schedule of RECIST assessments: “These assessments should occur irrespective of whether a subject is receiving study treatment or has previously discontinued study treatment for another discontinuation criterion and may have started alternative anticancer treatment.”	N	To clarify that it is not a requirement for subjects to have started alternative treatment.
Other	27-May-2020	Section 3.1 – Added sentence: “Investigators should notify the CRO again when progression is assessed at a subsequent timepoint.”	N	Added text to clarify (should have been included in CSP)
Other	27-May-2020	Section 3.1 – Updated to clarify that imaging assessments will be sent to AZ appointed CRO rather than central reader.	N	Clarification
Other	27-May-2020	Section 3.1.1 – Updated text regarding the purpose of the providing prior radiotherapy information to independent readers.	N	Clarification.
Other	27-May-2020	Section 3.1.1 – Updated text to “"Further details of the BICR are documented in the Independent Review Charter (IRC) and the CNS IRC."	N	To clarify that there are 2 charters.

Other	27-May-2020	Section 3.2.2.8, 3.2.2.9, 3.2.2.10, 3.2.2.11 Updated title to add "or death"	Y (v1.0)	Correction.
Other	27-May-2020	Section 3.2.2.8 – Removed underlined text: "Any subject not known to have died at the time of analysis and not known to have discontinued study treatment will be censored based on the last recorded date on which the subject was known to be alive (i.e. last known alive date taken from SURVIVE module)."	N/A	To ensure last known date alive used, regardless of module.
Other	27-May-2020	Section 3.2.2.9 – Added text to clarify that second progression must have occurred during or after treatment with subsequent anti-cancer treatment	N	Clarification.
Other	27-May-2020	Updated formatting and section headers throughout	N/A	Correction required.
Other	27-May-2020	Section 3.4.1.2 – Updated to remove reference to PhUSE guidelines.	N/A	Not relevant. Guidance is from TA SAP.
Other	27-May-2020	Section 3.6 – Added text "These will be formally analyzed in the CSR for all Part 2 screened subjects with evaluable results using the baseline result from Part 2 of screening, provided at least 15% of subjects are randomized under a local test."	N	Not many sites using pre-existing local results so numbers of subjects with both central and local tests could be small.
Other	27-May-2020	Section 4 – Added underlined text: "The primary endpoint of PFS and the secondary endpoint of CNS PFS will be tested only at the primary analysis when approximately 120 PFS events have been observed in 200 subjects (60% maturity).	Y (v1.0)	Clarification.
Other	27-May-2020	Section 4.2.9 Updated to use 'sd' rather than 'SD' to denote standard deviation.	N/A	'SD' already used to denote Stable Disease.

Other	27-May-2020	Section 4.2.10.2 – Updated to add underlined text: "Baseline tumor EGFR mutation status will be compared between tumor DNA and plasma ctDNA using the Kappa coefficient. Analysis will be carried out separately for each sensitizing mutation, Ex19Del and L858R, and as well as in aggregate."	N	For consistency with similar studies (FLAURA).
Other	27-May-2020	Section 4.2.10.2 – Added text: "The central **cobas®** result will be used where available, otherwise the local **cobas®** result will be used as reference."	N/A	To provide further details of how to handle cases where there is no central result to use as reference.
Other	27-May-2020	Section 4.2.10.2 – Amended "The baseline tumor EGFR mutation status will also be compared between the local and central cobas® EGFR mutation test v2 in subjects with evaluable results from baseline plasma samples." to "…baseline tumor samples"	N	Correction.
Other	27-May-2020	Section 4.2.10.2 – Updated text to clarify that two comparisons are carried out – tissue central cobas® vs plasma ctDNA test and tissue central cobas® vs local test.	N/A	Clarification.
Other	27-May-2020	Section 4.2.10.2 – Updated to clarify that only PPV applicable to 'central vs local' test comparison using central as reference. Other measures including PPV applicable to 'tissue vs plasma' comparison.	N/A	Clarification.
Other	27-May-2020	Section 3.6 Added further clarification that analysis only considered if at least 15% of subjects enrolled under local test.	N	Clarification.
Other	27-May-2020	Page 2. Phastar Study Statistician changed from PPD PPD to PPD	N/A	Change to Phastar Study Statistician
Other	27-May-2020	Section 1.2 Study Design. Updated version of CSP from 1.0 to 2.0.	N/A	To reference new version of CSP.

Other	27-May-2020	Section 1.2. Text updated to clarify that recruitment will be monitored to ensure "approximately" ≥60% of subjects entering the study have received prior concurrent chemoradiation.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Added text to state that after BICR-confirmed disease progression per RECIST v1.1, treatment assigned may be unblinded and subjects in both treatment arms may receive open-label osimertinib for as long as their treating physician considers them to be deriving clinical benefit. Also that subjects may not receive any other anti-cancer therapies between the discontinuation of study treatment and the start of open-label treatment with osimertinib.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated text to state that for subjects who have discontinued therapy due to disease progression, this must be BICR-confirmed as per RECIST v1.1.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Amended text from "Subjects who have discontinued treatment for any reason other than BICR-confirmed disease progression" to "Subjects who have discontinued treatment for any reason prior to BICR-confirmed disease progression".	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated text to clarify that AESIs will also be collected for 28 days following discontinuation, and for all events AEs/SAEs/AESIs, these should be followed-up until resolution.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated PFS occurrence of event estimates.	Y (2.0)	To align with CSP.

Other	27-May-2020	Section 1.2. Updated text to clarify that The study blind will be broken at the Sponsor level at the primary analysis of PFS, but sites and subjects will remain blind until after the final OS analysis.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Amended text from "After the final OS analysis, AstraZeneca will continue to supply open-label osimertinib to subjects until disease progression occurs as judged by the investigator, or until meeting any other discontinuation criteria" to "After the final OS analysis, AstraZeneca will continue to supply open-label osimertinib to subjects until disease progression occurs or unacceptable toxicity as judged by the investigator."	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2 Figure 1. Updated figure to align with CSP.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.3. Added underlined text "Approximately 200 subjects will be randomized, globally, in a 2:1 ratio"	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.3. In final two paragraphs clarified that the primary analysis of PFS and final analysis of OS in the China cohort will not take place until after the respective analyses have taken place in the global cohort.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 2.2. Added text to clarify that the content of the Important Protocol Deviations table is not an exhaustive list.	N	Client request. To avoid updates the SAP when further Important Protocol Deviations are added.

Other	27-May-2020	Section 2.2. Added underlined text "No locally advanced, unresectable (Stage III) NSCLC (V8 IASCL), except those with pre-existing local positive test results (Exon 19 or L858R)".	Y (V2.0)	To align with CSP.
Other	27-May-2020	Section 3.4.6. Added text to clarify that where appropriate if no digital ECG is captured prior to start of study treatment, but non-digital ECG is available, then this can be used as a baseline recoding.	Y (v2.0)	To align with CSP.
Other	27-May-2020	Section 4.2.5.1. Added underlined text "prior to chemoradiation strategy" and "entered into IVRS/IWRS".	Y (v2.0)	Clarification required.
Other	27-May-2020	Section 4.2.5.3. Added underlined text "entered into IVRS/IWRS".	Y (v2.0)	Clarification required.
Other	27-May-2020	Section 6. Removed changes of analysis from protocol as SAP now aligned with the protocol v2.0	Y (v2.0)	To align with CSP.
Other	27-May-2020	Throughout, updated "RECIST 1.1" to "RECIST v1.1".	Y (v2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Test updated to clarify that subjects assigned to osimertinib may continue to receive osimertinib 'in an open-label fashion'	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 2.1.2. SAS definition updated to include example of a subject receiving any active treatment	N/A	Clarification required.
Other	27-May-2020	Section 2.2. IPD text updated to refer to the Non-Compliance Handling Plan. 1st IPD removed as this is now covered by updates to 2nd IPD. Various updates to IPD description.	Y (2.0)	Client request,
Other	27-May-2020	Section 3.4.1.2. Updated text to clarify that the rules for missing end dates also apply for CM.	N/A	Correction.

Other	27-May-2020	Section 4.1. Added that SAS version 9.1 (as a minimum) will be used for all analyses	N/A	As per TA SAP.
Other	27-May-2020	Section 4.2.2. Clarified that the drug dictionary used to code medications received prior to, concomitantly or post treatment are to be coded by the WHO-DD.	N/A	Client request.
Other	27-May-2020	Section 4.2.4.2.1 Quantitative interactions section moved under section 4.2.5.3 Subgroup analyses of PFS as more relevant here.	N/A	Correction.
Other	27-May-2020	Section 4.2.7.1. 'Time to symptom improvement' updated to 'Symptom improvement rate'.	Y (v2.0)	Correction.
Other	27-May-2020	Section 4.2.10.2 Updated text to clarify analyses to be performed	N/A	Clarification
Other	27-May-2020	Section 1.2. Added text to state that after BICR-confirmed disease progression per RECIST v1.1, treatment assigned may be unblinded and subjects in both treatment arms may receive open-label osimertinib for as long as their treating physician considers them to be deriving clinical benefit. Also that subjects may not receive any other anti-cancer therapies between the discontinuation of study treatment and the start of open-label treatment with osimertinib.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated text to state that for subjects who have discontinued therapy due to disease progression, this must be BICR-confirmed as per RECIST v1.1.	Y (2.0)	To align with CSP.

Other	27-May-2020	Section 1.2. Amended text from "Subjects who have discontinued treatment for any reason other than BICR-confirmed disease progression" to "Subjects who have discontinued treatment for any reason prior to BICR-confirmed disease progression".	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated text to clarify that AESIs will also be collected for 28 days following discontinuation, and for all events AEs/SAEs/AESIs, these should be followed-up until resolution.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated PFS occurrence of event estimates.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Updated text to clarify that The study blind will be broken at the Sponsor level at the primary analysis of PFS, but sites and subjects will remain blind until after the final OS analysis.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Amended text from "After the final OS analysis, AstraZeneca will continue to supply open-label osimertinib to subjects until disease progression occurs as judged by the investigator, or until meeting any other discontinuation criteria" to "After the final OS analysis, AstraZeneca will continue to supply open-label osimertinib to subjects until disease progression occurs or unacceptable toxicity as judged by the investigator."	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.2 Figure 1. Updated figure to align with CSP.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 1.3. Added underlined text "Approximately 200 subjects will be randomized, <u>globally</u>, in a 2:1 ratio"	Y (2.0)	To align with CSP.

Other	27-May-2020	Section 1.3. In final two paragraphs clarified that the primary analysis of PFS and final analysis of OS in the China cohort will not take place until after the respective analyses have taken place in the global cohort.	Y (2.0)	To align with CSP.
Other	27-May-2020	Section 2.2. Added text to clarify that the content of the Important Protocol Deviations table is not an exhaustive list.	N	Client request. To avoid updates the SAP when further Important Protocol Deviations are added.
Other	27-May-2020	Section 2.2. Added underlined text "No locally advanced, unresectable (Stage III) NSCLC (V8 IASCL), except those with pre-existing local positive test results (Exon 19 or L858R)".	Y (V2.0)	To align with CSP.
Other	27-May-2020	Section 3.4.6. Added text to clarify that where appropriate if no digital ECG is captured prior to start of study treatment, but non-digital ECG is available, then this can be used as a baseline recoding.	Y (v2.0)	To align with CSP.
Other	27-May-2020	Section 4.2.5.1. Added underlined text "prior to chemoradiation strategy" and "entered into IVRS/IWRS".	Y (v2.0)	Clarification required.
Other	27-May-2020	Section 4.2.5.3. Added underlined text "entered into IVRS/IWRS".	Y (v2.0)	Clarification required.
Other	27-May-2020	Section 6. Removed changes of analysis from protocol as SAP now aligned with the protocol v2.0	Y (v2.0)	To align with CSP.
Other	27-May-2020	Throughout, updated "RECIST 1.1" to "RECIST v1.1".	Y (v2.0)	To align with CSP.
Other	27-May-2020	Section 1.2. Test updated to clarify that subjects assigned to osimertinib may continue to receive osimertinib 'in an open-label fashion'	Y (2.0)	To align with CSP.

Other	27-May-2020	Section 2.1.2. SAS definition updated to include example of a subject receiving any active treatment	N/A	Clarification required.
Other	27-May-2020	Section 2.2. IPD text updated to refer to the Non-Compliance Handling Plan. 1st IPD removed as this is now covered by updates to 2nd IPD. Various updates to IPD description.	Y (2.0)	Client request,
Other	27-May-2020	Section 3.4.1.2. Updated text to clarify that the rules for missing end dates also apply for CM.	N/A	Correction.
Other	27-May-2020	Section 4.1. Added that SAS version 9.1 (as a minimum) will be used for all analyses	N/A	As per TA SAP.
Other	27-May-2020	Section 4.2.2. Clarified that the drug dictionary used to code medications received prior to, concomitantly or post treatment are to be coded by the WHO-DD.	N/A	Client request.
Other	27-May-2020	Section 4.2.4.2.1 Quantitative interactions section moved under section 4.2.5.3 Subgroup analyses of PFS as more relevant here.	N/A	Correction.
Other	27-May-2020	Section 4.2.7.1. 'Time to symptom improvement' updated to 'Symptom improvement rate'.	Y (v2.0)	Correction.
Other	27-May-2020	Section 4.2.10.2 Updated text to clarify analyses to be performed	N/A	Clarification

* Pre-specified categories are:
Primary or secondary endpoints; Statistical analysis method for the primary or secondary endpoints; Derivation of primary or secondary endpoints; Multiple Testing Procedure; Data presentations; Other

1 STUDY DETAILS

1.1 Study objectives

1.1.1 Primary Objective

Table 1 Primary Objective

Primary Objective	Endpoint/Variable:
To assess the efficacy of osimertinib treatment compared with placebo as measured by progression-free survival (PFS)	• PFS using BICR assessment according to RECIST v1.1 • Sensitivity analysis of PFS using Investigator assessment according to RECIST v1.1
BICR=Blinded Independent Central Review; PFS = Progression-Free Survival; RECIST=Response Evaluation Criteria in Solid Tumors.	

1.1.2 Secondary Objectives

Table 2 Secondary Objectives

Secondary Objective	Endpoint/Variable:
To assess the efficacy of osimertinib treatment compared with placebo by assessment of PFS in subjects with: • EGFR Ex19del or L858R mutation • EGFR Ex19del or L858R mutation detectable in plasma-derived ctDNA	• PFS using BICR assessment according to RECIST v1.1 • Sensitivity analysis of PFS using Investigator assessment according to RECIST v1.1
To assess the efficacy of osimertinib versus placebo on CNS PFS	• Time to CNS PFS (time to the earliest of CNS progression or death) using BICR assessments according to RECIST v1.1 • Cumulative incidence rate of CNS PFS by BICR at 12 and 24 months
To further assess the efficacy of osimertinib compared with placebo	• OS • ORR • DoR • DCR and tumor shrinkage • TTDM All assessed by BICR according to RECIST v1.1 • TTD

To further assess the efficacy of osimertinib compared to placebo post progression	• PFS2 • TFST • TSST
To assess disease-related symptoms and health-related QoL in subjects treated with osimertinib compared with placebo.	• Change from baseline in EORTC QLQ-C30 • Change from baseline in EORTC QLQ-LC13
To assess the safety and tolerability profile of osimertinib compared with placebo	• AEs (graded by CTCAE v5) • Clinical chemistry, hematology and urinalysis • Vital signs (pulse and blood pressure), physical examination, weight • ECG parameters • LVEF • WHO Performance Status
To assess the PK of osimertinib	• Trough plasma concentrations of osimertinib, and its metabolite AZ5104 If conducted, PK Parameters ($CL_{ss/F}$, $C_{ss,min}$, $C_{ss,max}$, and AUC_{ss}) may be derived using population PK analysis and reported separately to the CSR. Data from this study may form part of a pooled analysis with data from other studies.

AE=Adverse Event; AUC_{ss}=Area Under Plasma Concentration Time Curve During any Dosing Interval at Steady State [amount time/volume]; BICR=Blinded Independent Central Review; CLss/F=Apparent Total Body Clearance at Steady State; CNS=Central Nervous System; $C_{ss,\,min}$=Minimum Plasma Concentration at Steady State; $C_{ss,max}$=Maximum Plasma Concentration at Steady State; ctDNA=circulating tumor DNA; DNA=Deoxyribonucleic Acid; DoR=Duration of Response; DCR=Disease Control Rate; ECG=Electrocardiogram; EGFR=Epidermal Growth Factor Receptor; EORTC QLQ-C30=European Organization for Research and Treatment of Cancer 30-Item Core Quality of Life Questionnaire; EORTC QLQ-LC13=European Organization for Research and Treatment of Cancer 13-Item Core Quality of Life Questionnaire; LVEF=Left Ventricular ejection fraction; ORR=Objective Response Rate; OS = Overall Survival; PFS = Progression-Free Survival; PFS2=Time from Randomization to Second Progression; PK=Pharmacokinetic(s); PRO=Patient Reported Outcomes; QoL=Quality of Life; RECIST=Response Evaluation Criteria in Solid Tumors; TTD= Time to Treatment Discontinuation; TTDM=Time To Death or Distant Metastases; TFST=Time to First Subsequent Therapy; TSST=Time to Second Subsequent Therapy.

1.1.3 Exploratory Objectives

Table 3 Exploratory Objectives

Exploratory Objectives	Endpoint/Variable:
To assess potential treatment-related adverse effects in subjects treated with osimertinib compared with placebo using PRO-CTCAE	The PRO-CTCAE questionnaire will be used to identify change in treatment-related symptoms

To assess the subjects' overall impression of the severity of their cancer symptoms using PGIS	PGIS: Proportion of subjects assessing current symptom severity.
To compare osimertinib treatment with placebo treatment on health state utility	The EQ-5D-5L health state utility index will be used to derive health state utility based on subject reported data.
To compare health resource use associated with osimertinib treatment versus placebo	HRU Module.
To investigate the relationship between osimertinib (and metabolite) PK and selected endpoints (which may include efficacy, safety and/or PRO), where deemed appropriate	Correlation of PK with other primary, secondary or exploratory endpoints in subjects treated with osimertinib.*
To compare the baseline tumor EGFR mutation status in screened subjects with evaluable results from baseline plasma samples. Data may be used to support diagnostic development.	Comparison of EGFR mutation status between tumor deoxyribonucleic acid (DNA) and plasma-derived ctDNA.
To compare the local EGFR mutation test result used for subject selection with the retrospective central **cobas®** EGFR Mutation Test v2 results from baseline tumor samples.	Comparison of EGFR mutation status between the local EGFR mutation test results and central **cobas®** EGFR Mutation Test v2 results from tumor samples with evaluable results.
To collect and store DNA for future exploratory research into genes/genetic variation that may influence PK or response to osimertinib (i.e. absorption, distribution, metabolism, excretion, safety and efficacy) and/or susceptibility to/development of cancers	Correlation of polymorphisms with variation in PK, pharmacodynamics, safety or response observed in subjects treated with osimertinib or comparator.*
To assess the relationship between PK and blood-borne biomarkers.	Correlation of blood-based biomarkers, including alterations in ctDNA, with variation in PK*.
To collect and store tumor samples to evaluate the association between exploratory biomarkers and key efficacy endpoints	Key markers to include, but not limited to mutations, amplifications, or expression changes in EGFR mutations, HER2 and MET.*
To collect and store plasma for isolation of ctDNA and to evaluate the association between exploratory biomarkers and key efficacy endpoints	Longitudinal analysis of ctDNA for mutations, amplifications or expression changes in EGFR, HER2, MET and other genes. *

To collect and store plasma to assess the relationship between blood borne biomarkers and key efficacy endpoints	Biomarkers will include but will not be limited to growth factors, or cytokines.*
To identify innate resistance mechanisms to study treatment.	Assessment of innate resistance mechanisms including but not limited to identify mutations, amplifications, or expression changes in EGFR, HER2 and MET in baseline ctDNA and/or tissue biopsies.*
To identify acquired resistance mechanisms to study treatment.	Assessment of resistance mechanisms including but not limited to identify mutations, amplifications, or expression changes in EGFR, HER2 and MET in ctDNA and/or tissue biopsies taken at time of disease progression.*
To conduct exploratory research on tissue and plasma samples into factors that may influence susceptibility to/development of NSCLC/cancer and/or response to osimertinib (where response is defined broadly to include efficacy, tolerability or safety). Tissue and plasma samples may be used to support diagnostic development.	Samples may be analyzed retrospectively for novel biomarker discovery research and/or diagnostic development.*
*Samples may be analyzed retrospectively. The results of this research will be reported either in the CSR itself or as an addendum, or separately in a scientific report or publication. The results of this research may be pooled with data from other studies with the study drug to generate hypotheses to be tested in future research.	
ctDNA=circulating tumor DNA; DNA=Deoxyribonucleic Acid; EGFR=Epidermal Growth Factor Receptor; MET=tyrosine-protein kinase Met; CSR=Clinical Study Report; CTCAE=Common Terminology Criteria for Adverse Events; EQ-5D-5L=EuroQoL 5-Dimension 5-Levels; HER2=Human Epidermal Growth Factor Receptor 2; NSCLC=non-small cell lung cancer; PGIS=Patients Global Impression of Severity; PRO-CTCAE=Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Events.	

1.2 Study design

This statistical analysis plan (SAP) contains a more detailed description of the analyses in the clinical study protocol (CSP) and is based on version 5.0 of the CSP.

This is a Phase III double-blind, randomized, placebo-controlled, multicenter international study assessing the efficacy and safety of osimertinib as a maintenance therapy in subjects with locally advanced unresectable (Stage III), non-small cell lung cancer (NSCLC) with epidermal growth factor receptor (EGFR) mutations (Ex19Del and L858R), whose disease has not progressed following definitive platinum based chemoradiation therapy. A schematic diagram of the overall study design is shown in Figure 1.

Study entry is permitted based on the detection of an Ex19del and/or L858R mutation via central, tissue-based EGFR testing using the **cobas®** EGFR Mutation Test v2, or from a pre-existing local EGFR test result obtained using a tissue-based FDA-approved CDx for

TAGRISSO (i.e., **cobas**® EGFR Mutation Test v2 or FoundationOne® CDx test), performed in a CLIA-certified (USA sites) or an accredited local laboratory (sites outside of the USA). Subjects without a local positive test result will be required to undergo a two-part screening process. In Part I screening, subjects will be asked to consent to provide their recent archival tumor tissue sample for prospective EGFR mutation testing. Part II screening will be conducted following completion of chemoradiation and will be carried out in subjects who have passed Part I screening. In addition, subjects with a local positive result can enter Part II screening directly after completion of chemoradiation. For these subjects, investigators will be asked to provide an archival formalin-fixed and paraffin-embedded (FFPE) tumor tissue sample, where available, for retrospective central analysis of EGFR mutation status in Part II screening for comparison with the local test. However, subject enrollment will be based upon the local test.

Following completion of chemoradiation, all subjects will be required to have a baseline computerized tomography (CT) or magnetic resonance imaging scan (MRI) scan and contrast-enhanced MRI of the brain within 28 days prior to randomization. Subjects whose disease has not progressed are eligible for the study.

To reflect global clinical practice, recruitment will be monitored on an ongoing basis and will be managed to ensure that the majority (approximately ≥60%) of subjects entering the study have received prior concurrent chemoradiation (CCRT).

Randomization must occur ≤6 weeks (i.e. 42 days with no visit windowing applied) following completion of chemoradiation. Subjects will be stratified by prior chemoradiation strategy (CCRT vs SCRT), tumor stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity) to allow separate randomization for China.

Subjects who meet all the inclusion criteria and none of the exclusion criteria for this study will be randomized 2:1 to receive either osimertinib or matching placebo 80mg po QD until objective radiological disease progression per RECIST v1.1 which is confirmed by BICR or until another treatment discontinuation criterion is met.

Treatment assignment may be unblinded for each subject with BICR-confirmed disease progression per RECIST v1.1. Subjects assigned to osimertinib may continue to receive osimertinib in an open-label fashion if, in the opinion of their treating physician, they are continuing to derive clinical benefit. Subjects assigned to placebo may receive open-label osimertinib, in accordance with local clinical practice and the judgement of their treating physician. Subjects must not receive any other anti-cancer therapies between discontinuation of study treatment and the start of treatment with open-label osimertinib. For all subjects, post-progression treatment with osimertinib may continue as long as the treating physician considers the subject to be deriving clinical benefit.

Tumor assessments will be performed using (i) CT scan (preferred) or MRI each preferably with contrast, of the chest and abdomen (including liver and adrenal glands) and additional anatomy imaging, as indicated by signs and symptoms of the subject; (ii) contrast enhanced T1w MRI of the brain. These will take place every 8 weeks (relative to randomization) until 48 weeks, then every 12 weeks thereafter, until objective radiological disease progression occurs and is confirmed by BICR.

Subjects who have discontinued study treatment due to BICR-confirmed disease progression per RECIST v1.1 will enter survival follow-up in which survival status, health resource use module, subsequent anti-cancer treatment and second progression date will be recorded every 12 weeks until death, withdrawal of consent or the end of study. Subject reported outcomes will be assessed at weeks 8, 16 and 32 post discontinuation. Subjects who have discontinued treatment for any reason prior to BICR-confirmed disease progression will enter progression follow-up and should continue to have tumor assessments according to the schedule of assessments until disease progression confirmed by BICR.

Subjects will undergo safety assessments at baseline, Week 2, Week 4 and every 4 weeks until Visit 9/Week 24, every 8 weeks until Visit 12/Week 48, then every 12 weeks afterwards until treatment discontinuation.

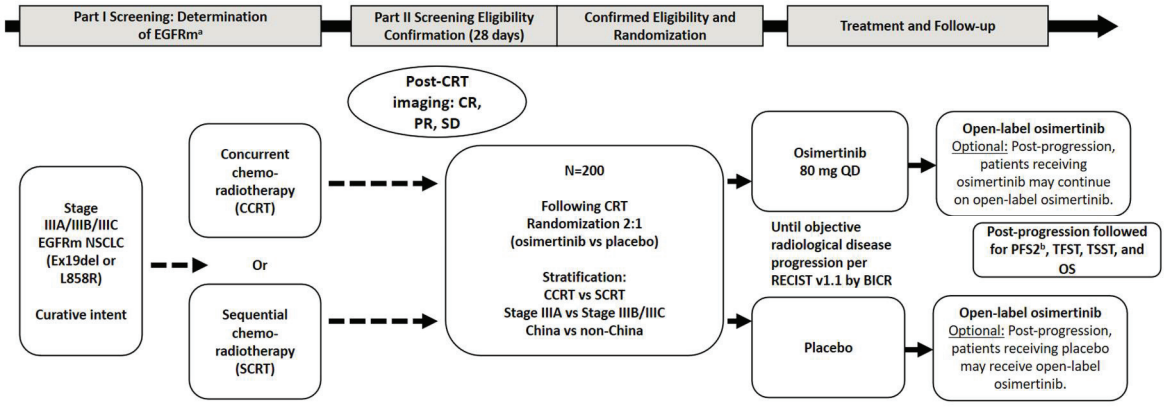
Safety assessments will include clinical chemistry, hematology, urinalysis, ECG and LVEF measurement with ECHO/MUGA scans. Tumor and blood sampling for biomarker analysis will be performed. PK sampling will also be performed. PROs and Health Resource Use (HRU) for monitoring subject's health condition and resource use respectively will be collected during the study. Following discontinuation of study drug, a treatment discontinuation visit will be performed. AEs/SAEs/AESIs will be collected for 28 days following discontinuation and followed-up until resolution.

Subjects will be followed as per the study protocol until data cut-off for the primary analysis when approximately 120 PFS BICR events have been observed. This is estimated to occur approximately 48 months following first subject randomized, based on a 40-month recruitment period. The analysis of PFS in the China cohort will be conducted at the same time as the analysis of PFS in the overall (global) population.

Following the primary PFS analysis, subjects who are continuing to receive study treatment will have study assessments as per the schedule of assessments for subjects receiving study treatment prior to the primary PFS analysis with the exception of tumor assessments. Following the primary PFS analysis, subjects will have tumor assessment in accordance with the local clinical practice and formal RECIST measurements will not be collected. The study blind will be broken (at the Sponsor level) at the primary analysis of PFS. Sites and subjects will remain blinded until study completion i.e. after final OS analysis.

The final analysis of OS will be conducted at approximately 60% maturity when approximately 120 death events (across both arms) have occurred. The final analysis of OS in the China cohort will be conducted at the same time as the final analysis of OS for the overall population. After the final OS analysis, AstraZeneca will continue to supply open-label osimertinib until the subject stops deriving clinical benefit (as judged by the investigator), or until osimertinib is commercially available for use in the first-line setting in the subjects' respective country/territory.

Figure 1 Overall study design



[a] Subjects with a positive local EGFR test result via a tissue-based FDA approved CDx for TAGRISSO (i.e., cobas® EGFR Mutation Test v2 or FoundationOne® CDx test) do not need to undergo Part I screening.
b Assessment of PFS2 will not be collected after the primary PFS analysis

1.3 Number of subjects

This study will enroll approximately 1333 subjects to account for those who do not have EGFR Ex19del or L858R mutation detected, sample attrition and an 85% screen fail rate for other reasons, to randomize approximately 200 subjects.

Approximately 200 subjects will be randomized, globally, in a 2:1 ratio (osimertinib to placebo). Of those, it is planned that approximately 30 to 40 patients will be recruited in China. This is being done to ensure adequate Chinese subject participation to satisfy China Regulatory Authority requirements. The China cohort will support standalone safety and efficacy analyses of subjects from China.

Subjects randomized will have achieved a complete response (CR), partial response (PR), or have stable disease (SD) following definitive, platinum-based, chemoradiation as evaluated by the investigator according to RECIST v1.1. Randomization will be stratified by: prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity).

In order to reflect global clinical practice, recruitment will be monitored on an ongoing basis and will be managed to ensure that the majority (approximately ≥60%) of subjects entering the study have received prior CCRT. Study entry is permitted based on detection of Ex19del and/or L858R mutation via central, tissue-based EGFR testing using the **cobas**® EGFR Mutation Test v2, or from a pre-existing local EGFR test result obtained using a tissue-based FDA-approved CDx for TAGRISSO (i.e., **cobas**® EGFR Mutation Test v2 or FoundationOne® CDx test), performed in a CLIA-certified (USA sites) or an accredited local laboratory (sites outside of the USA).

The primary analysis of PFS in the global cohort and analysis of PFS in the China cohort will occur when approximately 120 progression events have been observed from the globally randomized subjects. With 120 PFS BICR events the study has 90% power to show a statistically significant difference in PFS at the 2-sided 5% level if the assumed true treatment effect is a HR of 0.53 for the comparison of osimertinib versus placebo, assuming a median of 8 months PFS for placebo. The smallest treatment difference that could be statistically significant being a HR of 0.68.

2 ANALYSIS SETS

2.1 Definition of analysis sets

Four analysis sets are defined for this study. Table 4 gives a summary of outcome variables and their analysis sets.

2.1.1 Full analysis set

The full analysis set (FAS) will include all randomized subjects (i.e. the Intention-To-Treat [ITT] population). The FAS will be used for the analysis of the primary objective and all other efficacy analyses. Treatment groups will be compared by randomized study treatment, regardless of the treatment actually received i.e. subjects who were randomized but did not go on to receive study treatment are included in the analysis in the treatment group to which they were randomized.

2.1.2 Safety analysis set

The safety analysis set (SAF) will consist of all randomized subjects who received at least one dose of study treatment. Safety data will not be formally analyzed but summarized using the safety analysis set, according to the treatment received; i.e. erroneously treated subjects (e.g. those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received. If a subject receives any active treatment then they will be summarized in the active treatment arm.

2.1.3 Pharmacokinetic analysis set

The pharmacokinetic (PK) analysis set is defined as subjects in the SAF who have at least one measurable PK concentration, supported by the relevant date and time of this sample; and for each time a PK sample was taken, the dosing data for that day; and for samples taken after multiple dosing, the dosing data for 7 continuous day of osimertinib dosing

prior to the sample day as well as the sample day. For any individual sample to be included in the PK analysis set, the full sample data and dosing data need to be present for that sample.

2.1.4 Evaluable for response analysis set (subset of Full Analysis Set)

The evaluable for response analysis set (EFR) will be all subjects in the FAS who have measurable disease at baseline according to the BICR of baseline imaging data.

The primary analysis of PFS and ORR by BICR in the FAS will be repeated using the evaluable for response analysis set as sensitivity analyses.

Table 4 Summary of outcome variables and analysis set

Outcome variable	Analysis set
Demography data	Full analysis set
PK data	PK analysis set
Efficacy data	
PFS*, CNS PFS*, OS, ORR*, DoR*, DCR*, TTDM*, Tumor shrinkage,*, PFS2, TFST, TSST, TTD, PROs.	Full analysis set
PFS*, and ORR* (sensitivity analyses)	Evaluable for response analysis set (subset of Full analysis set)
Safety data	
Exposure, AEs, Laboratory measurements, Vital signs, WHO, ECG and Left ventricular ejection fraction.	Safety analysis set
*Using BICR assessment as per RECIST 1.1.	
PK=Pharmacokinetic(s); PFS=Progression-Free Survival; CNS PFS= Central Nervous System Progression-Free Survival; OS=Overall Survival; ORR=Objective Response Rate; DoR=Duration of Response; DCR=Disease Control Rate; TTDM=Time to Death or Distant Metastases; PFS2=Time from randomization to Second Progression on subsequent treatment; TFST=Time to First Subsequent Therapy; TSST=Time to Second Subsequent Therapy; TTD=Time to Treatment Discontinuation; PROs=Patient Reported Outcomes; RECIST=Response Evaluation Criteria in Solid Tumors; AEs=Adverse Events; WHO=World Health Organization; ECG= Electrocardiogram.	

2.2 Protocol deviations

The following general categories will be considered important protocol deviations that may significantly impact the reliability of the study data or may significantly affect the overall interpretation of the primary and/or secondary study results, some of which will be programmatically derived from the eCRF data, and all will be listed in the CSR. A full list of IPDs and descriptions can be found in the Protocol Deviations (PD) Plan.

Criteria type	Important Protocol Deviation Description
Inclusion	- No locally advanced, unresectable (Stage III) NSCLC (V8 IASCL) - No confirmation that the tumor harbors one of the two common EGFR mutations known to be associated with EGFR-TKI sensitivity (Ex19del, L858R), either alone or in combination with other EGFR mutations, assessed by a tissue-based FDA-approved CDx for TAGRISSO (i.e., cobas® EGFR Mutation Test v2 or FoundationOne® CDx test), performed in a CLIA certified (USA sites) or an accredited local laboratory (sites outside of the USA) or by central testing (using the cobas® EGFR Mutation Test v2). - Subject has not received either concurrent chemoradiation or sequential chemoradiation regimens as defined in CSP - Subject has had disease progression during or following definitive platinum based, chemoradiation therapy - Subject has not received platinum-based chemotherapy regimen containing one of the following agents: etoposide, vinblastine, vinorelbine, paclitaxel, docetaxel, or pemetrexed, according to the local standard of care regimens Subject has received gemcitabine along with radiation. Note: Gemcitabine is permitted if used prior to radiation but not with radiation.
Exclusion	- History of interstitial lung disease (ILD) prior to chemoradiation - Symptomatic pneumonitis following chemoradiation. - Prior treatment with any prior chemotherapy, radiation therapy, immunotherapy or investigational agents for NSCLC outside of that received in the definitive setting for Stage III disease. - Prior treatment with EGFR-TKI therapy. - The laboratory values demonstrate that subject has inadequate bone marrow reserve or organ function (CTCAE ≥ Grade 2), as per CSP exclusion criterion 6.

IP Administration/Study treatment	- Subject received incorrect study treatment to that to which they were randomized. - Subjects who were randomized but did not receive IP - Subject restarted study treatment after experiencing G3 or higher ILD; G2 ILD not resolved within 4 weeks or recurrent symptomatic ILD following prior dose interruption and IP re-challenge. - Subject restarted study treatment after QTc interval prolongation with signs/symptoms of serious arrhythmia.
Disallowed medications	- Subject receives other anticancer agents (including traditional Chinese medicine registered for anti-cancer effects), investigational agents, radiotherapy (for reasons other than locally advanced NSCLC).
Procedure/tests	- Baseline tumor assessment (RECIST v1.1) and brain MRI performed more than 42 days before randomization. - RECIST scan performed outside of the scheduled window on more than two occasions throughout the study in one subject. Note: Scans OOW do not have to be consecutive. - Missing any RECIST assessment for efficacy.

Subjects who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 2.1. During the study, decisions on how to handle errors in treatment dispensing (with regard to continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

3 PRIMARY, SECONDARY, AND EXPLORATORY VARIABLES

3.1 Derivation of RECIST visit response

The analysis of the primary endpoint PFS and the analyses of the secondary endpoints, ORR, DoR, DCR, TTDM and tumor shrinkage will be based on BICR assessments using RECIST v1.1. A sensitivity analysis of PFS will be based on the site investigator assessments according to RECIST v1.1.

Subjects with measurable disease and/or non-measurable and/or no evidence of disease assessed at baseline by CT/ MRI will be entered in this study.

CT (preferred) or MRI of chest and abdomen and MRI of the brain will be performed at all tumor imaging visits. The baseline assessment is part of the screening procedures and should be performed ≤28 days prior to randomization. The imaging modality used for baseline tumor assessment, CT/MRI for chest and abdomen and MRI for brain, should be kept the same consistently at each subsequent follow-up assessment throughout the study if possible.

Efficacy for all subjects will be assessed by objective tumor assessments every 8 weeks (relative to the date of randomization) until 48 weeks, then every 12 weeks thereafter, until objective radiological disease progression as defined by RECIST v1.1 and as confirmed by BICR. These assessments should occur irrespective of whether a subject is receiving study treatment or has previously discontinued study treatment for another discontinuation criterion and may have started alternative anticancer treatment. Additional scans should be performed if clinically indicated, i.e. if disease progression is suspected. An MRI scan with contrast should be performed in the event of suspected CNS progression. If an unscheduled assessment is performed, and the subject's disease has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits.

All imaging assessments (including unscheduled visit scans, radiotherapy planning scan, and high-resolution CT) will be sent to the AstraZeneca-appointed Clinical Research Organization (CRO) on an ongoing basis. Assessment of scans by BICR will be triggered only upon investigator-assessed progression and the results of the BICR will be reported back promptly to sites. If the BICR confirms disease progression, study treatment will be discontinued, and the subjects will enter survival follow-up. If the BICR does not confirm disease progression, study treatment and tumor assessments should be continued in line with the schedule of assessments. Investigators should notify the CRO again when progression is assessed at a subsequent timepoint. Since the primary analysis of the study is based on BICR, it is important that study treatment and scheduled imaging assessments continue until progression confirmed by BICR. Progression includes disease recurrence for subjects with no evidence of disease at baseline.

At the time of the primary PFS analysis, all scans for all subjects will have BICR review. If at this time it is identified that a subject has disease progression by BICR that has not been identified by the investigator, AstraZeneca will contact the investigator to discuss whether continued treatment with study medication is appropriate.

3.1.1 Blinded Independent Central Review (BICR) Assessment using RECIST v1.1

All imaging scans for tumor assessments, including unscheduled visit scans, should be duplicated and collected on an ongoing basis and sent to the appointed CRO to enable blinded independent central review (BICR). Information about prior radiotherapy from the CRF (if available) will also be provided to the BICR CRO to inform the independent readers of the anatomical region of previous definitive radiotherapy, to allow the selection of appropriate target and non-target lesions (if present at baseline), and to designate new

lesions at follow-up that are outside the Lung or Regional Lymph Node area as ‘distant’ metastases for derivation of TTDM.

The imaging scans will be reviewed by two independent radiologists using RECIST v1.1 and will be adjudicated, if required (i.e. two reviewers review the scans and adjudication is performed by a separate reviewer in case of a disagreement). For each subject, the BICR will define the overall visit response (i.e. the response obtained overall at each visit by assessing TLs, NTLs and new lesions) data and no programmatic derivation of visit response is necessary. (For subjects with TLs at baseline: CR, PR, SD, PD, NE; for subjects with NTLs only: CR, SD, PD, NE; for subjects with no disease identified at baseline: PD, no evidence of disease [NED], NE). If a subject has had a tumor assessment that cannot be evaluated, then the subject will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in target lesion dimensions. RECIST v1.1 criteria will be used to assess each subject's tumor response to treatment and allow calculation of PFS, ORR, DoR, DCR, tumor shrinkage and TTDM.

Following BICR-confirmed progression, until the time of primary analysis, subjects should have tumor assessments as per standard local practice for assessment of second progression on a subsequent treatment (PFS2). These local-practice scans should not be sent to the appointed CRO.

3.1.1.1 CNS BICR

For the purpose of assessing CNS PFS, all subjects will have a baseline brain MRI scan. The CNS BICR assessment is separate from the whole-body BICR assessment and is comprised of independent neuroradiologists review.

As all subjects are expected not to have any extra-thoracic metastases at baseline, PFS CNS will be assessed based on the appearance of new CNS lesions. However, if the neuroradiologist identifies CNS lesions at baseline then the overall response at each visit will be assessed using modified RECIST guidelines.

Further details of the BICR are documented in the Independent Review Charter (IRC) and CNS IRC.

3.1.2 Site Investigator Assessment Using RECIST v1.1

For all subjects the RECIST v1.1 tumor response data will be used to determine each subject's visit response.

At each visit, subjects will be programmatically assigned a RECIST v1.1 visit response of CR, PR, SD, NE, NED or PD depending on the status of their disease compared with baseline and previous assessments. Baseline will be assessed within the 28 days prior to randomization. If a subject has had a tumor assessment that cannot be evaluated, then the subject will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD).

3.1.2.1 Target lesions (TLs)

A measurable lesion is one which can be accurately measured at baseline as ≥10mm in the longest diameter (except lymph nodes which must have short axis diameter of ≥15mm) with CT or MRI and which is suitable for accurate repeated measurements.

A maximum of 5 measurable lesions (with a maximum of 2 lesions per organ), representative of all lesions involved suitable for accurate repeated measurement, should be identified as TLs at baseline. Lymph nodes, in any location, are collectively considered as a single organ, with a maximum of 2 lymph nodes as TLs. A bilateral organ (e.g. adrenal glands) is considered as a single organ. Each segmented organ (e.g. liver) or lobular organ (e.g. lung) is considered as a single organ.

Prior irradiated lesions may be considered measurable and selected as TLs provided they fulfil the other criteria for measurability.

Note: For subjects who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see Section 3.1.3 for further details). If a subject does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

For subjects with no disease at baseline (i.e. no TLs and no NTLs), evaluation of overall visit responses will be based on absence/presence of new lesions. If no TLs and no NTLs are recorded at a visit, both the TL and NTL visit response will be recorded as NA and the overall visit response will be no evidence of disease (NED). If a new lesion is observed, then the overall visit response will be PD.

Table 5 provides the definitions of the criteria used to determine objective tumor visit response for TL.

Table 5 TL visit responses (RECIST v1.1)

Complete Response (CR)	Disappearance of all TLs since baseline. 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 the diameters of TL, taking as reference the baseline sum of diameters.
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.
Progressive Disease (PD)	At least a 20% increase in the sum of diameters of TLs, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5mm.
Not Evaluable (NE)	Only relevant if any of the TLs at follow-up were not assessed or not evaluable (e.g. missing anatomy) or had a lesion intervention at this visit. Note: if the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
CR=Complete Response; PR=Partial Response; PD=Progressive Disease; NE=Not Evaluable; SD=Stable Disease; TL=Target Lesion.	

Rounding of TL data

For calculation of PD and PR for TLs, percentage changes from baseline and previous minimum should be rounded to one decimal place 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
- An 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" (the previous minimum sum of diameters) even assuming the non-recorded TLs have disappeared

Note: the nadir can only be taken from assessments where all the TLs had a longest diameter recorded.

Lymph nodes

For lymph nodes, if the size reduces to <10mm 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 <10mm and all other TLs are 0mm, although the sum may be >0mm, the calculation of TL response should be overwritten 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. 0mm 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. 0mm 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 and the sum of lesions meets the criteria for PD 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. 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 5mm will be entered into the database and used in TL calculations. However, a smaller value may be used if the 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, then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

According to the study design, it is possible that subjects have some residual disease that has been irradiated prior to their study entry. Lesions within sites of prior radiotherapy

may be considered measurable and selected as a target lesion providing they fulfil the other criteria for measurability.

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 and 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 tumors:

- 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, the lesion diameter (for those lesions with intervention) will be treated as missing and if ≤1/3 of the TLs have missing measurements then the results will be scaled up as described in the 'Scaling' section below. If the scaling results in a visit response of PD then the subject 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. Subjects with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention have a value of 0 (or <10mm 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 intervention) 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 ≥5mm from nadir).

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; it had a baseline measure of 29.3cm. The sum of lesions 1-4 at the follow-up is 26 cm. The sum of the corresponding lesions at the nadir visit is 26.8 cm. Scale up as follows to give an estimated TL sum of 28.4cm:

$$\frac{26}{26.8} \times 29.3 = 28.4cm$$

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 0mm.

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.

3.1.2.2 Non-target lesions (NTLs) and New lesions

All other lesions (or sites of disease) not recorded as TL should be identified as NTL at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits. At each visit an overall assessment of the NTL response should be recorded by the Investigator. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit (see Table 6).

Table 6 NTL visit response (RECIST v1.1)

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10mm 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 (NN)	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 subjects 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.

CR=Complete Response; PD=Progressive Disease; NN=Non-Complete Response/Non-Progressive Disease; NE=Not Evaluable; NA=Not Applicable; NTLs=Non-Target Lesions; TLs=Target Lesions.

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 tumor 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.

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

Subjects with ‘symptomatic progression’ requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumor assessments where possible until objective disease progression is observed.

3.1.2.3 New lesions

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.

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 RECIST v1.1 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 tumor. If a new lesion is equivocal, for example because of its small size, the treatment and tumor assessments should be continued until the previously new lesion has been assessed as unequivocal and then the progression date should be declared using the date of the initial scan when the new lesion first appeared.

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.

3.1.3 Overall visit response

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 lesions	Non-target lesions	New lesions	Overall Response
CR	CR	No	CR
CR	NA	No	CR
CR	Non-CR/Non-PD	No	PR
CR	NE	No	PR
PR	Non-PD or NE	No	PR
SD	Non-PD or NE	No	SD
NE	Non-PD or NE	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD
NA	CR	No	CR
NA	Non-CR/Non-PD	No	SD
NA	NE	No	NE
NA	NA	No	NED

CR=Complete Response; NA=Not Applicable; PD=Progressive Disease; Non-CR= Non-Complete Response; Non-PD=Non-Progressive Disease; PR=Partial Response; SD=Stable Disease; NE=Not Evaluable; NED=No Evidence of Disease.

3.2 Efficacy outcome variables

3.2.1 Primary outcome: progression-free survival (PFS)

PFS (per RECIST v1.1 as assessed by BICR) is defined as the time from randomization until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy prior to progression (i.e. date of PFS event or

censoring – date of randomization + 1). Subjects 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 non-missing RECIST assessment. However, if the subject progresses or dies after two or more missed visits, the subject will be censored at the time of the latest non-missing RECIST v1.1 assessment prior to the two missed visits (note: NE visit is not considered a missed visit).

Given the scheduled visit assessment scheme (i.e. 8-weekly for the first 48 weeks and then 12-weekly thereafter) the definition of 2 missed visits will be as follows:

- If the previous RECIST v1.1 assessment is before Week 40 (study day 274) then two missed visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks)
- If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from eight-weekly to twelve-weekly this will equate to 22 weeks (i.e. take the average of 8 and 12 weeks which gives 10 weeks and then apply same rationale, hence 2 x 10 weeks + 1 week for an early assessment + 1 week for a late assessment = 22 weeks). The time period for the previous RECIST assessment will be from study days 274 to 329 (i.e. week 39 to week 47)
- From week 47 onwards (study day 330 when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 26 weeks (i.e. 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks)

If the subject has no post-baseline assessments or does not have baseline data, they will be censored at Day 1 unless they die within two visits of baseline (16 weeks plus 1 week allowing for late assessments i.e. 119 days).

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 selected single reviewer described in 3.1.1
- For investigator 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 investigator assessments, when censoring a subject for PFS the subject 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 Secondary outcomes

3.2.2.1 CNS progression-free survival (CNS PFS)

CNS PFS (per RECIST v1.1 as assessed by BICR in a separate CNS charter) is defined as the time from randomization until the date of CNS objective disease progression or death (by any cause in the absence of CNS progression) regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy prior to CNS progression (i.e. date of CNS PFS event or censoring – date of randomization + 1). Subjects who do not have CNS progression nor have died at the time of analysis will be censored at the time of their last non-missing RECIST assessment. However, if the subject progresses in the CNS or dies after two or more missed visits, as defined in the PFS analysis, the subject will be censored at the time of the latest non-missing RECIST v1.1 assessment prior to the two missed visits.

3.2.2.2 Overall survival (OS)

OS is defined as the time from the date of randomization until death due to any cause (i.e. date of death or censoring – date of randomization + 1). Any subject not known to have died at the time of analysis will be censored based on the last recorded date on which the subject was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

If a subject 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

b. For Missing day and month – using the 1st of January

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.

Note: Survival calls will be made within one week following the date of data cut-off for the primary PFS analysis and the final OS analysis. If subjects are confirmed to be alive or if the death date is post the data cut-off date, these subjects will be censored at the date of data cut-off. Death dates may be found by checking publicly available death registries (as applicable under local laws). The date of contact will not be used.

3.2.2.3 Objective response rate (ORR)

ORR is defined as the percentage of subjects with at least one BICR-assessed visit response of CR or PR. The denominator is defined as the subset of all randomized subjects.

A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 28 days after the visit when the response was

first observed with no evidence of progression between the initial and CR/PR confirmation visit.

Data obtained up until progression, or last non-missing assessment in the absence of progression, will be included in the assessment of ORR except when subjects who discontinue randomized treatment without progression, receive a subsequent anti-cancer therapy and then respond; they will not be included as responders in the ORR.

3.2.2.4 Duration of response (DoR)

DoR will be 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) using BICR assessments. The end of response should coincide with the date of progression or death from any cause used for the RECIST v1.1 PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of PR or CR. If a subject does not progress following a response, then their DoR will use the PFS censoring time. DoR will not be defined for those subjects who do not have documented response.

3.2.2.5 Disease control rate (DCR)

Disease control rate is defined to be the percentage of subjects who have a best objective response (BoR) of CR or PR or SD at ≥8 weeks, prior to any PD event. The 8-week time point will allow for a visit window and be defined as on or after study day 49 (allowing for a 1-week early assessment).

3.2.2.6 Best Objective Response (BoR)

BoR is calculated based on the overall visit responses from each BICR RECIST assessment. It is the best response a subject has had following randomization up to and including RECIST progression or the last non-missing assessment in the absence of RECIST progression.

BoR will be determined programmatically based on RECIST from the overall visit response using all BICR data up until the first progression event or subsequent anti-cancer therapy, whichever is the earliest. BoR will be based on RECIST using the following response categories and in the following order:

- Response with categories: CR (confirmed), PR (confirmed)
- Non-response with categories:
    - SD with categories:
        - unconfirmed CR
        - unconfirmed PR
        - SD (>=8 weeks)
    - NE with categories:
        - SD (<8 weeks)
        - no baseline RECIST
        - no post baseline RECIST assessments in the absence of death

- Progression with categories:
  - RECIST progressions and death in the absence of RECIST progression

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 8 weeks minus 1 week, i.e. at least 49 days (to allow for an early assessment within the assessment window), after randomization. For CR/PR, the initial overall visit assessment that showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

For subjects who die with no RECIST assessments, if the death occurs ≤17 weeks (i.e. 16 weeks + 1 weeks to allow for a late assessment within the 2 assessment windows) after randomization, then BoR will be assigned to the progression (PD) category. For subjects who die with no non-missing RECIST assessments, if the death occurs >17 weeks after randomization then BoR will be assigned to the NE category.

3.2.2.7 Depth of response

Depth of response (or tumor shrinkage or change in tumor size) will be assessed using the BICR RECIST v1.1 tumor responses in target lesions. The absolute change and percentage change from baseline in sum of tumor size at each assessment will be calculated. The best change in tumor size (i.e. depth of response) is the largest decrease from baseline or the smallest increase from baseline in the absence of a reduction and will include all assessments prior to progression or start of subsequent anti-cancer therapy. If best percentage change cannot be calculated due to missing data (including if the subject has no TLs at baseline), a value of +20% will be imputed as the best percentage change in the following situations (otherwise best percentage change will be left as missing):

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

3.2.2.8 Time to death or distant metastases (TTDM)

Time to death or distant metastases (TTDM) will be defined as the time from the date of randomization until the first date of distant metastasis or date of death in the absence of distant metastasis. Distant metastasis is defined as any new lesion that is detected on a scan that is anywhere other than Lung or Regional Lymph Node according to RECIST v1.1 or proven by biopsy. Subjects who have not developed distant metastasis or died at the time of analysis will be censored at the time of the latest date of assessment from their last non-missing RECIST v1.1 assessment.

However, if the subject has distant metastasis or dies after 2 or more missed visits, the subject will be censored at the time of the latest non-missing RECIST v1.1 assessment

prior to the 2 missed visits. The 2 missed visit rule will be as detailed in Section 3.2.1 for the primary analysis of PFS. If the subject has no post-baseline assessments or does not have baseline data, he/she will be censored at Day 1 (date of randomization) unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for late assessments within the two visit windows i.e. 119 days).

3.2.2.9 Time to study treatment discontinuation or death (TTD)

Time to study treatment discontinuation or death (TTD) is defined as the time from randomization to the earlier of the date of study treatment discontinuation (regardless of the reason for study treatment discontinuation) or death (i.e. date of study treatment discontinuation/death or censoring – date of randomization + 1). Any subject not known to have died at the time of analysis and not known to have discontinued study treatment will be censored based on the last recorded date on which the subject was known to be alive. Subjects who were randomized but did not receive treatment will be censored at the date of randomization.

3.2.2.10 Time from randomization to second progression-free survival 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 progression event following first objective disease progression, subsequent to the first subsequent therapy, or death. The first objective progression includes progression occurring after 2 missed visits (i.e., date of PFS2 event or censoring – date of randomization + 1) and will be based on Investigator assessment. Following a first progression as assessed by Investigator, subjects will have their progression status recorded every 12 weeks per local standard clinical practice to assess time to second progression (PFS2) on a subsequent treatment. A subject's second progression status is defined according to the local practice and may involve any of: objective radiological progression (preferred), symptomatic progression, or death. Scans will be performed according to the local practice and formal RECIST measurements will not be collected for assessment of PFS2. The date of PFS2 assessment and Investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the source documents and the eCRF.

Subjects alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without a second disease progression; i.e. censored at the last progression assessment date or latest PFS2 assessment date if the subject has not had a second progression or death.

3.2.2.11 Time to first subsequent therapy or death (TFST)

Time to first subsequent therapy or death (TFST) is defined as the time from the date of randomization to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomized treatment, or death (i.e. date of first subsequent cancer therapy/death or censoring – date of randomization + 1). Any subject not known to have had a first subsequent anti-cancer therapy will be censored at the last date that the subject

was known not to have received a first subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a subject terminated the study for reason other than death before first subsequent therapy, these subjects will be censored at the earliest of their last known to be alive and termination dates. Subjects not receiving randomized treatment would have TFST calculated as time from date of randomization to the initial therapy or death.

3.2.2.12 Time to second subsequent therapy or death (TSST)

Time to second subsequent therapy or death (TSST) is defined as the time from the date of randomization to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of randomized treatment, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomization + 1). Any subject not known to have had a second anti-cancer subsequent therapy will be censored at the last date that the subject was known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a subject terminated the study for reason other than death before second subsequent therapy, these subjects will be censored at the earliest of their last known to be alive and termination dates. Subjects not receiving randomized treatment would have TSST calculated in the same way, i.e. time from date of randomization to the subsequent therapy or death.

3.3 Patient-reported outcomes (PROs)

PROs will be assessed using the EORTC QLQ-C30, EORTC QLQ-LC13, PRO CTCAE, PGIS and EQ-5D-5L (used as a health utility measure). All items/questionnaires will be scored according to published scoring guidelines. All PRO analyses will be based on the FAS, expect for PRO-CTCAE which will use the SAF, and will take place at the time of the primary PFS analysis. PRO data will not be collected for any subject after the primary PFS analysis. Endpoints based on EORTC QLC-C30 and EORTC QLC-LC13 scales are secondary outcomes; all other PRO endpoints are exploratory outcomes. PROs will be analyzed using the PRO data collected at the planned timepoints according to the Schedule of Activities in the CSP.

3.3.1 EORTC QLQ-C30 and EORTC QLQ-LC13

The EORTC QLQ-C30 consists of 30 items and measures cancer subjects' functioning (HRQoL) and symptoms (Aaronson et al 1993). Questions can be grouped into 5 multi-item functional scales (physical, role, emotional, cognitive, and social); 3 multi-item symptom scales (fatigue, pain, nausea/vomiting); a 2-item global health status / QoL scale; 5 single items assessing additional symptoms commonly reported by cancer subjects (dyspnea, loss of appetite, insomnia, constipation, diarrhea) and 1 item on the financial impact of the disease.

The EORTC QLQ-LC13 includes questions assessing symptoms: cough, hemoptysis, dyspnea, site specific pain (pain in chest, pain in arm or shoulder, pain in other part), and treatment-related side effects: 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.

An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales, each of the functional scales, and the global health status scale according to the EORTC QLQ-C30 Scoring Manual and EORTC QLQ-LC13 instructions. Higher scores on the global health status and functional scales indicate better health status/function, whereas higher scores on symptom scales represent greater symptom severity.

For each symptom/scale, if less than 50% of the subscale items are missing, the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscale. If at least 50% of the items are missing, that subscale also will be treated as missing. The dyspnea scale from the EORTC QLQ-LC13 will only be used if all 3 items have been scored otherwise, the items will be treated as single-item measures.

Definition of compliance and evaluability rates

Compliance with the EORTC QLQ-C30 and EORTC QLQ-LC13 will be calculated, separately for each questionnaire:

$$\text{Compliance rate} = \frac{\text{number of evaluable forms}}{\text{number of expected forms}} \times 100$$

Evaluability rates for the EORTC QLQ-C30 and EORTC QLQ-LC13 will also be calculated, separately for each questionnaire:

$$\text{Evaluability rate} = \frac{\text{number of evaluable forms}}{\text{number of received forms}} \times 100$$

Where an expected, evaluable, and received questionnaire is defined as follow:

- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time, i.e. a questionnaire from a subject who has not withdrawn from the study at the scheduled assessment time but excluding subjects in countries with no available translation
- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing
- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed

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 (Osoba 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 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, or worsening as shown in Table 8. Subjects with no baseline data will be excluded from analyses.

Table 8 Visit responses for symptoms and HRQoL

Score	Change from Baseline	Visit Response
QLQ-C30 symptom scales/items	≥+10 ≤-10 Otherwise	Worsened Improved Stable
QLQ-LC13 symptom scales/items	≥+10 ≤-10 Otherwise	Worsened Improved Stable
QLQ-C30 functional scales and global health status/QoL	≥+10 ≤-10 Otherwise	Improved Worsened Stable

QoL Quality of Life; QLQ C30 30-Item Quality-of-Life Questionnaire; QLQ-LC13 13-Item Lung Cancer Quality-of-Life Questionnaire.

Time to symptom deterioration (QLQ-C30 and QLQ-LC13)

For each of the symptoms scales in the EORTC QLQ-C30 and EORTC QLQ-LC13, and for subjects with a baseline score <=90, time to symptom deterioration will be defined as the time from randomization until the date of the first confirmed clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) that is confirmed at the next consecutive non-missing assessment (except if it was the subject's last available assessment) or death (by any cause) in the absence of a clinically meaningful confirmed symptom deterioration, regardless of whether the subject withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration. Death will be included as an event only if the death occurs within two visits of the last PRO assessment where the symptom change could be evaluated. All symptoms, functioning subscales, and the GHS/QoL subscale from the QLQ-C30 and dyspnea, cough, and chest pain from QLQ-LC13 will be included in the analysis.

Subjects whose symptoms (as measured by EORTC QLQ-C30 and EORTC QLQ-LC13) have not shown a confirmed clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if a confirmed symptom deterioration occurs after two or more missed PRO assessment visits or the subject dies after two or more missed PRO assessment visits, the subject will be censored at the time of the last PRO assessment where the symptom could be evaluated (prior to the two missed assessment visits).

Following the approach for the primary analysis 2 missed visits will be defined as follows:

EORTC QLQ-C30

- If previous assessment is <Week 4 (study day 27) then 2 missed visits = 60 days (2 x 4 week + 2 days for early assessment + 2 days for late assessment).
- If 2 missed visits occur over the period when the scheduled frequency of EORTC QLQ-C30 assessments changes from 4-weekly to 8-weekly this will equate to 89 days (i.e. take the average of 4 and 8 weeks which gives 42 days, 2*42 days + 2 days for early assessment + 3 days for late assessment). The time period for the previous EORTC QLQ-C30 assessment will be from study day 27 to 53.
- From Week 8 onwards (study day 54 when the scheduling changes to 8-weekly assessments) two missing visits will equate to 118 days (i.e. 2 x 8 weeks + 3 days for an early assessment + 3 days for a late assessment).

EORTC QLQ-LC13

- If previous assessment is <Week 7 (study day 48) then 2 missed visits = 18 days (2 x 1 week + 2 days for early assessment + 2 days for late assessment).
- If 2 missed visits occur over the period when the scheduled frequency of EORTC QLQ-LC13 assessments changes from weekly to 4-weekly this will equate to 40 days (i.e. take the average of 1 and 4 weeks which gives 18 days (taking a conservative approach), 2*18 days + 2 days for early assessment + 3 days for late assessment). The time period for the previous EORTC QLQ-LC13 assessment will be from study day 48 to 53.
- From Week 8 onwards (study day 54 when the scheduling changes to 4-weekly assessments) two missing visits will equate to 62 days (i.e. 2 x 4 weeks + 3 days for an early assessment + 3 days for a late assessment.

If a subject has no evaluable visits they will be censored at day 1. The population for the analysis of time to symptom deterioration will be a subset of the FAS who have non-missing baseline scores of ≤90.

Time to HRQoL/function deterioration (QLQ-C30)

For the functional scales or the global health status/QoL in QLQ-C30 and for subjects with a baseline score >=10, time to deterioration will be defined as the time from the date of randomization until the date of the first confirmed clinically meaningful deterioration (a decrease in the function scales or the global health status/QoL from baseline of ≥10) that is confirmed at the next consecutive non-missing visit (except if it was the subject's last available assessment) or death (by any cause) in the absence of a clinically meaningful confirmed deterioration, regardless of whether the subject withdraws from study treatment or receives another anticancer therapy prior to HRQoL/function deterioration. Death will

be included as an event only if the death occurs within 2 visits of the last PRO assessment where the HRQoL/function change could be evaluated.

Subjects whose HRQoL (as measured by the GHS/QoL subscale from the EORTC QLQ-C30) have not shown a confirmed clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the HRQoL/function could be evaluated. Also, if a confirmed HRQoL deterioration occurs after two or more missed PRO assessment visits or the subject dies after 2 or more missed PRO assessment visits, the subject will be censored at the time of the last PRO assessment where HRQoL/function could be evaluated following the same approach as for EORTC QLQ-C30 symptom deterioration.

If a subject has no evaluable visits they will be censored at day 1. The population for the analysis of time to HRQoL/function deterioration will include a subset of the FAS who have non-missing baseline scores of ≥10.

Symptom improvement rate (QLQ-C30 and QLQ-LC13)

The symptom improvement rate will be defined as the number (%) of subjects with two consecutive assessments at least 21 days apart which showed a clinically meaningful improvement (a decrease from baseline score ≥10 for EORTC QLQ-LC13 and EORTC QLQ-C30 scales) in that symptom from baseline. The denominator will consist of a subset of the FAS who have a non-missing baseline symptom score ≥10.

3.3.2 Patients Global Impression of Severity (PGIS)

The PGIS item is included to assess how a subject perceives his/her overall current severity of cancer symptoms. Subjects will choose from response options from 1. no symptoms, 2. very mild, 3. mild, 4. moderate, 5. severe and 6. very severe

3.3.3 Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Event (PRO-CTCAE)

The PRO-CTCAE system was developed by the NCI in recognition that collecting potential treatment-related symptoms directly from subjects can improve the accuracy and efficiency of reporting of adverse events. The PRO-CTCAE will only be administered in those countries where a linguistically validated version exists. The PRO-CTCAE is an item-bank of symptoms experienced by subjects while undergoing treatment of their cancer. To date, 81 symptoms of the CTCAE have been identified to be amenable to subject reporting but not all items are administered in any one clinical trial. Response options cover frequency, severity, and interference with usual activities. For this study, 14 items are considered relevant for this cancer treatment.

3.3.4 EuroQoL 5-Dimension 5-Levels (EQ-5D-5L)

The EuroQoL 5-Dimension (EQ-5D) is a standardized measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal (EuroQol Group 1990). The EQ-5D-5L questionnaire assesses 5

dimensions as follows: mobility, selfcare, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 response options ("no problems," "slight problems," "moderate problems," "severe problems," and "extreme problems") that reflect increasing levels of difficulty (EuroQol Group 2015).

The subject will be asked to indicate his/her current health state by selecting the most appropriate level in each of the 5 dimensions. In addition to the descriptive system, subjects 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.

3.3.5 Health resource use module

Healthcare Resource Use Module will be completed by the investigational site for any healthcare resource use between visits. Up to the point of primary PFS analysis, the site will ask subjects for any health resource use between visits (i.e. excluding routine follow-up clinic visits associated with the clinical trial but including both planned and unplanned admissions) at study visits as per Table 1 of the CSP.

Healthcare resource use related to treatment and the underlying disease will be recorded specifically hospital episodes and symptoms for admission.

3.4 Safety assessments

Safety and tolerability will be assessed by adverse events, laboratory data, physical examinations, vital signs, ECG, LVEF by echocardiogram/MUGA scans and WHO performance status.

3.4.1 General considerations for safety assessments

3.4.1.1 Safety Follow-up

Total Safety Follow-up = min((last dose date + 28 days), date of withdrawal of consent, date of death, date of DCO, date of first dose of subsequent anti-cancer therapy [excluding palliative radiotherapy]) – first dose date +1.

3.4.1.2 Handling missing data

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.

Imputation of missing dates

For AE and concomitant medications imputation methods will be used to assess if an observation with a partially missing or completely missing start or end date is treatment emergent.

Subjects with a partial date of birth (i.e., for those countries where year of birth only is given) will have the 1st of the month imputed if the date is missing, and 1st Jan imputed if the date and month is missing.

Concomitant medication, radiotherapy and adverse event start dates will be imputed as follows:

- If year is 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
- If year is present and month and day are missing, then impute January 1st unless year is the same as first dose date then impute first dose date.
- If year and month are present and day is missing, impute 1st of the month unless the month is the same month of first dose of study drug then impute first dose date.

For missing end CM, radiotherapy and AE dates, the following will be applied:

- Missing day - Impute the last day of the month unless month is the same as month of the last dose of study drug then impute last dose date. For prior anti-cancer medications impute date of informed consent if month is the same as month informed consent was provided.

- Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date. For prior anti-cancer medications impute date of informed consent if month is the same as month informed consent was provided.

- Completely Missing – consider when it started in relation to study drug. If the AE/medication has stopped and start date is prior to first dose date then impute the 1st dose date, if it started on or after first dose date then impute a date that is after the last dose date (i.e. last dose + 1), unless this is for a prior anti-cancer medication then impute the date of informed consent.

  For partial subsequent anti-cancer therapy dates, the following rules will be applied for missing start dates:

  · Missing day: If the month is the same as treatment end date then impute to the day after treatment, otherwise first day of the month.

  · Missing day and month: If year is the same as treatment end date then impute to the day after treatment, otherwise 1st January of the same year as anti-cancer therapy date.

3.4.2 Exposure and dose interruptions

Exposure (i.e. duration of treatment) will be defined as follows:

Total (or intended) exposure of osimertinib or placebo:

- Total (or intended) exposure (months) = (last dose date where dose > 0 mg – first dose date + 1)/(30.4375).

Actual exposure of osimertinib or placebo:

- Actual exposure (months) = ((last dose date where dose > 0 mg – first dose date + 1) – total duration of dose interruptions)/(30.4375), where a dose interruption is defined as any length of time where the subject has not taken any of the planned daily dose.

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

Missed or forgotten doses

Missed and forgotten doses should be recorded on the EX module as a dose interruption with the reason recorded as "Subject forgot to take dose". These missed or forgotten doses will not be included as dose interruptions in the summary tables, but the information will appear in the listing for dosing. However, these missed and forgotten doses will be considered in the derivation of actual exposure.

Subjects who permanently discontinue study treatment during a dose interruption
If a subject 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.

3.4.3 Laboratory assessments

Laboratory data will be collected throughout the study, from screening to follow-up visits as described in Table 1 of the CSP. The clinical chemistry, hematology and urinalysis will be collected as described in Section 8.2.1 of the CSP.

3.4.4 Physical examinations

A complete physical examination will be performed and include an assessment of the following: general appearance, respiratory, cardiovascular, abdomen and additional systems as deemed clinically appropriate (e.g. skin). Height will be measured only during Part II screening. Physical examination and weight will be performed at timelines as specified in Table 1 of the CSP.

3.4.5 Vital signs

The vital signs (pulse and blood pressure) will be evaluated according to the assessment schedules in Table 1 of the CSP.

3.4.6 Electrocardiograms (ECGs)

ECGs will be performed at the visits specified in Table 1 of the CSP. In addition to this, they should be performed in the event of any cardiac AE.

ECG data will be collected digitally and will be transferred electronically for central analysis as described in the study specific ECG manual. If no digital ECG is captured prior to the start of study treatment, where an appropriate non-digital ECG is available, the non-digital ECG will be used as a baseline recording.
The mean of the available ECG results at each timepoint will be derived for the analysis.

Heart rate, PR, R-R, QRS, QT intervals and QTcF will be determined and reviewed where Fridericia's formula is applied to calculate the corrected QT interval:

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

where QT and RR are in seconds.

3.4.7 Left ventricular ejection fraction (LVEF)

An echocardiogram or MUGA scan to assess LVEF will be performed at the visits as shown in Table 1 in the CSP.

3.4.8 WHO performance status

WHO performance status will be assessed at the scheduled visits as indicated in Table 1 of the CSP according to WHO criteria.

3.4.9 Adverse events (AEs)

AEs will be collected throughout the study from the time of signature of Screening Part II ICF throughout the treatment period until the 28-day follow-up visit (i.e. 28 days after last dose of study drug). In addition, any AEs related to study procedures occurring during Screening Part I will be collected. Any AE occurring before treatment with IP, or after the start or subsequent cancer therapy will be included in the data listings but will not be included in the summary tables of AEs. Serious adverse events (SAEs) will be recorded during the study from time of signing the first ICF. Adverse Events of Special Interest (AESIs) will be recorded from the time of signing the Screening Part II ICF through to post the final OS analysis if applicable. See protocol Section 8.3.2 for further details.

The Medical Dictionary for Regulatory Activities (MedDRA) dictionary (using the latest or current MedDRA version) will be used to code the AEs to a preferred term. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 5).

COVID-19 related AEs

Confirmed or Suspected COVID-19 AEs are defined as all AEs occurring during the pandemic timeframe with a preferred term within the AE search criteria developed by the latest MedDRA Maintenance and Support Services Organization (MSSO) guidance for COVID-19.

COVID-19 associated AEs are defined as all confirmed and suspected COVID-19 AEs defined above, plus all other AEs occurring within <7 days before and <30 days after the start date of all the confirmed COVID-19 AEs.

AEs of Special Interest (AESIs)

Adverse events of special interest (AESIs) are events of scientific and medical interest specific to the further understanding of osimertinib's safety profile and require close monitoring and rapid communication by the Investigator to the Sponsor.

An AESI may be serious or non-serious. The rapid reporting of these AESIs allows ongoing analysis of these events in order to characterize and understand them in association with the use of the IP.

The AESIs identified in CSP v5.0 for this study are pneumonitis, interstitial lung disease and radiation pneumonitis. Cardiac effects is identified as a programme-wide AESI and is also monitored.

3.5 Pharmacokinetics

Pharmacokinetic concentration data will be collected as per the protocol at pre-dose on Day 29, Day 85 and Day 169.

The PK samples may be subjected to analyses by AstraZeneca to investigate the presence and/or identity of additional drug metabolites and correlate PK with other primary, secondary, and exploratory endpoints in subjects treated with osimertinib. Analysis of samples from China will be performed as per local regulations. Full details of the analytical method used will be described in a separate bioanalytical report. Any results from such analyses will be reported separately from the CSR.

3.6 Other exploratory variables

Other exploratory data detailed in Section 1.1.3 will be collected for analysis. However, the results from this research will be reported either in the CSR itself or as an addendum, or in a scientific report or publication separately from the CSR.

The baseline tumor EGFR mutation status will be compared between tumor deoxyribonucleic acid (DNA) and plasma-derived circulating tumor DNA (ctDNA). Likewise, the baseline EGFR mutation status will be compared between the local and central **cobas®** EGFR mutation test v2 results. These analyses will be considered for inclusion in the CSR for all Part 2 screened subjects with evaluable results using the

baseline result from Part 2 of screening, provided at least 15% of subjects are randomized under a local test.

This includes investigating correlation of blood-based biomarkers, assessing innate resistance mechanisms to treatment and further PK analysis with correlation of polymorphisms with variation in PK, pharmacodynamics, safety or response in subjects who received osimertinib. The results from this research will be reported either in the CSR itself or as an addendum, or separately in a scientific report or publication. The results of this research may be pooled with data from other studies with the study drug to generate hypotheses to be tested in future research.

4 ANALYSIS METHODS

The formal statistical analyses will be performed to test the main hypothesis:

- H_0: No difference between osimertinib and placebo
- H_1: Difference between osimertinib and placebo

The primary objective of this study is to assess the efficacy of osimertinib compared with placebo in terms of PFS as assessed by BICR. Key secondary endpoints are OS and CNS PFS by BICR.

In order to strongly control the type I error at 5% 2-sided significance level, a sequential testing procedure will be implemented and after the testing of PFS, OS and CNS PFS will be tested (in that order). If any previous analysis in the sequence is not statistically significant, the alpha will not be transferred to subsequent analyses.

Multiple testing strategy

The multiple testing procedure (MTP) will define which significance levels should be applied to the interpretation of the raw p-values for the primary endpoint of PFS and key secondary endpoints, OS and CNS PFS. The family-wise error rate is strongly controlled at 5% (two sided) for these endpoints.

To provide strong control of the type I error rate, the primary endpoint of PFS and secondary endpoints, OS and CNS PFS, will be tested in sequential order. If any previous analysis in the sequence is not statistically significant, the alpha cannot be transferred to subsequent analyses. Hypotheses will be tested using a multiple testing procedure with an alpha recycling strategy (Burman et al 2009). With this approach, hypotheses will be tested in the predefined order of PFS, OS and CNS PFS.

The analyses of PFS, OS and CNS PFS endpoints will occur at the time of the primary analysis when approximately 120 PFS events by BICR assessment have been observed in approximately 200 subjects (approximately 60% maturity). The secondary endpoint of OS will be tested at both primary analysis (when analysing the primary endpoint of PFS) and final analysis of OS. The final analysis of OS will occur when approximately 120 death events have been reported in approximately 200 subjects (approximately 60% maturity).

The alpha (2-sided 5%) level allocated to OS will be controlled at the primary and final analyses by using the Lan-DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the alpha level applied at the primary analysis depends upon the information fraction. This approach will maintain an overall 2-sided 5% type I error across the two planned analyses of OS.

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 events at the final analysis time-point. For example, assuming a median OS on the placebo arm of 40 months and a median OS of 50 months on the osimertinib arm, 48 OS events were observed at the interim analysis, the information fraction would be 0.40 (48/120 events) for the interim analysis and 1.00 for the final analysis. This would result in a significance level for the interim analysis of 0.0008 (2-sided) and a significance level for the final analysis of 0.0497 (2-sided).

Any non-statistically significant OS analyses at the time of the primary analysis of PFS will not preclude further testing of OS.

If the OS analysis is statistically significant at the time of the PFS analysis or the final OS analysis, then the significance testing of CNS PFS will be performed at the full α=0.05 significance level (two-sided). If the OS analysis is not statistically significant at the time of the PFS analysis or the final OS analysis then the statistical significance testing of CNS PFS in will be not be performed.

4.1 General principles

The below mentioned general principles will be followed throughout the study:

- 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. 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 arm
- 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 9.1 (as a minimum) will be used for all analyses.
- Missing coding terms should be listed and summarized as 'Uncoded'

- A month is defined as 30.4375 days

In general, for efficacy and PRO endpoints, the last observed measurement before 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. All efficacy analysis will be performed on the FAS unless otherwise stated.

Safety data will be summarized and analyzed on the SAF and the last observation prior to first dose will be considered the baseline. In all summaries change from baseline 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 x 100.

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.

4.1.1 Visit windows

Time windows will need defining for any presentations that summarize values by visit. The following conventions should apply:

- The time windows should be exhaustive such that data recorded at any time point has the potential to be summarized. Inclusion within the time window should be based on the actual date and not the intended date of visit
- All unscheduled visit data should have the potential to be included in the summaries
- The window for visits following baseline will 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 consecutive visits then the upper limit will be taken as the midpoint value minus 1 day
- 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 subject
- For visit-based summaries:
    - If there is more than one value per subject within a time window then the closest value to the scheduled visit date should be summarized, or the earlier, in the event the values are equidistant from the nominal visit date. If there are two equally eligible assessments at a visit, then the worst value will be taken as the post-baseline value. The listings should highlight the

value for the subject 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 plots being produced that contain many cells with meaningless data, for each treatment group, visit data will only be summarized if the number of observations is greater than the minimum of 5

- To prevent summaries being produced with low precision, for each treatment group, visit data will only be summarized if the number of observations is ≥5. For summaries at a subject level, all values should be included, regardless of whether they appear in a corresponding visit-based summary, when deriving a subject level statistic such as a maximum
- If two visits are equally eligible to assess subject 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 can 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, the best value would be taken as baseline 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 data are summarized over time, study day will be calculated in relation to date of first treatment (safety endpoints) or date of randomization (efficacy endpoints)
- For PRO endpoints, baseline assessments must be completed prior to first dose of study medication and study day will be calculated relative to randomization

4.2 Analysis methods

4.2.1 Demographic and other baseline characteristics

Demographic and baseline subject characteristics will be listed and summarized for the FAS. Standard descriptive statistics will be presented for the continuous variables of:

- Age (years)
- Weight (kg)
- Height (cm)
- Body mass index (kg/m2), calculated as: $\frac{weight}{height^2}$
- Nicotine consumption (number of pack years)

- Mean lung radiotherapy dose (Gy) and v5 (%), v20 (%), Cardiac v30 (%), v45 (%) and v50 (%)
- Time from unresectable Stage III diagnosis to randomization (days)

The total counts and percentages of subjects will be presented for the categorical variables of:

- Age group (years) (grouped as <50, ≥50-<65, ≥65-<75, ≥75)
- Sex
- Race
- Ethnic group
- Disease Stage
- Overall disease classification
- Extent of disease
- Baseline TL size by BICR (mean(SD), median, and categories: <40, 40-79, 80-119, ≥120mm)
- Baseline TL size by Investigator (mean(SD), median, and categories: <40, 40-79, 80-119, ≥120mm)
- Histology type
- Smoking status (never, current, former)
- WHO performance status (0/1)
- Response to prior chemoradiation (CR, PR, SD, NE)
- PET Scan performed prior to chemoradiotherapy
- EGFR tissue mutation type (Ex19Del or L858R) from tests used to confirm study eligibility
- Previous disease-related treatment modalities
- Disease related medical history
- Surgical history

A listing of all subjects showing all the EGFR mutations identified by the **cobas**® central test will be presented. The number of subjects recruited in each country and each center will be presented by treatment group and in total. The number and percentage of subjects' potential prognostic factors and stratification factors recorded at randomization by IVRS will be presented. Listings of medical and surgical history will be produced.

4.2.2 Protocol Deviations

A summary table will be produced showing the number and percentage of subjects with any IPD and by category of IPD, which will include the individual IPDs as detailed in Section 2.2. In addition, the number and percentage of subjects with at least one COVID-19 pandemic related IPD will be summarized.

The individual subject data for IPDs is also listed.

In addition, each stratification factor as recorded by the IVRS and as recorded according to the eCRF will be cross-tabulated to evaluate any mis-stratifications.

In addition to the summary of the important protocol deviations above, other study deviations captured from the CRF module for inclusion/exclusion criteria will listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

4.2.3 Concomitant and other treatments

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 the eCRF.

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

Medications received prior to, concomitantly, or post-treatment will be coded using the World Health Organization Drug Dictionary (WHO-DD) encoding (using the latest or current WHO-DD version). Concomitant medications will be summarized for the FAS by ATC classification codes.

For the purpose of inclusion in prior and/or concomitant medication or therapy summaries, incomplete medication or radiotherapy start and stop dates will be imputed as detailed in Section 3.4.1.2.

Prior medications, concomitant and post-study 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 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 for the full analysis set.

The following summaries will be produced:

- Summary of prior medications (including radiotherapy type)
- Summary of concomitant medications
- Summary of post study treatment anti-cancer therapies

All concomitant medications and other treatment data will be listed.

Missing coding terms should be listed and summarized as “Uncoded”.

4.2.4 Exposure

Exposure will be summarized for the safety analysis set. The following summaries will be produced:

- Summary of duration of exposure of study treatment and of open-label osimertinib
- Summary of interruptions and reductions of study treatment

4.2.5 Subject disposition and data sets analyzed

Subject disposition will be listed and summarized for the FAS. Summaries will include the number and percentage of subjects:

- Enrolled (informed consent received)
- Randomized
- Treated
- Subjects ongoing study treatment and on-going study at the data cut-off
- Included in each analysis set (FAS, Safety, PK and evaluable for response)

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

4.2.6 COVID-19 Impact

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 subjects affected by the COVID-19 pandemic
- Listing for subjects with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic.

4.2.7 Primary outcome: PFS

4.2.7.1 Primary analysis of PFS

The analysis of PFS will be based on the BICR RECIST v1.1 assessments and will use a stratified log-rank test stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation strategy (IIIA vs IIIB/IIIC), and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity) for generation of the p-value.

The stratification factors will be based on the values entered into IVRS/IWRS at randomization, even if it is subsequently discovered that these values were incorrect.

The effect of osimertinib vs placebo will be estimated by a HR and 2-sided 95% CI which 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 stratified log-rank test statistic obtained from the LIFETEST procedure (with a TEST statement).

The assumption of proportionality will be assessed. Proportionality will be tested firstly by examining the plots of complementary log-log (event times) vs log (time) and, if necessary, a time dependent covariate will be fitted 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. If the resulting strata are too small (i.e. <10 events) the strata will be collapsed in the following pre-defined order to allow analysis. The China cohort strata will be collapsed first, followed by prior chemoradiation strategy and finally stage prior to chemoradiation. In the event of non-proportionality, the HR will be interpreted as an average HR over the observed extent of follow-up.

Kaplan-Meier (KM) plots will be presented by treatment arm. Summaries of the number and percentage of subjects experiencing a PFS event and the type of event (PD assessed by RECIST v1.1 or death) will be provided along with estimates and 95% CI for PFS rates at 6 monthly intervals and median PFS for each treatment using Kaplan-Meier technique. The treatment status at progression of subjects at the time of analysis will be summarized. This will include the number (%) of subjects who were on treatment at the time of progression, the number (%) of subjects who discontinued study treatment prior to progression, the number (%) of subjects who have not progressed and were on treatment or discontinued treatment and the number of subjects whose PFS event was death.

In addition, a summary of new lesions will be produced by treatment arm. A listing of all tumor assessment based on investigator and central review will also be produced.

4.2.7.2 Sensitivity analyses of PFS

4.2.7.2.1 Stratification according to the eCRF

In the event that there are any mis-stratifications during randomization, the stratified log rank test will be repeated on PFS, where the stratification factors are as recorded according to the eCRF. The HR and CI will be presented from the same method described as for the primary analysis described in Section 4.2.7.1.

4.2.7.2.2 Ascertainment bias

The possibility of bias in assessment and measurement of PFS by BICR will be assessed using the investigator assessment of disease progression by RECIST v1.1.

The stratified log-rank test will be repeated using investigator-assessed RECIST data to programmatically derive PFS. A Kaplan-Meier (KM) plot will be presented by treatment arm. Summaries of the number and percentage of subjects experiencing a PFS event and

the type of event (PD assessed by RECIST v1.1 or death) will be provided along with estimates and 95% CI for PFS rated at 6 monthly intervals and median PFS for each treatment using Kaplan-Meier technique. The HR and 95% CI will be presented.

If there is an important discrepancy between the primary analysis using the BICR data and this sensitivity analysis using the investigator data, then the proportion of subjects with BICR but no investigator progression will be summarized; such subjects have the potential to induce bias in the central review due to informative censoring. Disagreements between BICR and investigator reviews of RECIST progression will be presented for each treatment group. The summary will also include assessing ascertainment bias using the early discrepancy rate (EDR) and late discrepancy rate (LDR) (Amit, O. et al 2011). The EDR represents the positive predictive value of BICR assessment and quantifies the frequency with which the BICR declare progression early relative to Investigator within each treatment arm as a proportion of the total number of BICR assessed progressions. The LDR quantified the frequency that the BICR declare progression later than Investigator as a proportion of the total number of discrepancies within the treatment arm. If the distribution of discrepancies is similar between the treatment arms then this suggests the absence of evaluation bias favoring a particular treatment arm.

The EDR and LDR are calculated as follows:

$$EDR = \frac{\begin{array}{c}number\ of\ times\ BICR\ declares\ PD\ when\ Inv.\ does\ not\ + \\ number\ of\ times\ BICR\ declares\ PD\ \geq 2\ weeks\ earlier\ than\ Inv.\end{array}}{\begin{array}{c}number\ of\ times\ both\ BICR\ and\ Inv.\ declare\ PD \\ +\ number\ of\ times\ BICR\ declares\ PD\ when\ Inv.\ does\ not\end{array}}$$

$$LDR = \frac{\begin{array}{c}number\ of\ times\ Inv.\ declares\ PD\ when\ BICR\ does\ not\ + \\ number\ of\ times\ BICR\ declares\ PD \geq 2\ weeks\ later\ than\ Inv.\end{array}}{\begin{array}{c}number\ of\ times\ BICR\ declares\ PD\ when\ Inv.\ does\ not \\ +\ number\ of\ times\ Inv.\ declares\ PD\ when\ BICR\ does\ not \\ +\ number\ of\ times\ BICR\ declares\ PD \geq 2\ weeks\ later\ than\ Inv. \\ +\ number\ of\ times\ BICR\ declares\ PD \geq 2\ weeks\ earlier\ than\ Inv.\end{array}}$$

The EDR and LDR will be calculated for each treatment arm and the differential discordance around each measure will be defined as the rate on the experimental arm minus the rate on the control arm. A negative differential discordance for the EDR and/or positive differential discordance for the LDR are suggestive of a bias in the BICR assessment favoring the experimental arm.

4.2.7.2.3 Evaluation-time bias

In order to assess possible evaluation-time bias that could occur if scans are not performed at the protocol-scheduled time points, the midpoint between the time of progression and the previous non-missing RECIST assessment will be analyzed using a log rank test stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject

declaring themselves of non-Chinese ethnicity), as described for the primary analysis of PFS.

For subjects who die in the absence of progression, the date of death will be used to derive the PFS time used in the analysis. For those subjects not experiencing disease progression during the study the PFS time will be right-censored at the date of the last non-missing RECIST assessment. If the subject has no evaluable visits or does not have baseline data, they will be censored at Day 1. Note that midpoint values resulting in non-integer values should be rounded down. 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 subject-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR RECIST assessments.

4.2.7.2.4 Attrition bias

Possible attrition bias will be assessed by repeating the primary PFS analysis, except that the actual PFS event times, rather than the censored times, of subjects who progressed or died in the absence of progression immediately following two, or more, non-missing tumor assessments will be included. In addition, and within the same sensitivity analysis, subjects who take subsequent therapy prior to progression or death will be censored at their last non-missing assessment prior to taking the subsequent therapy.

This analysis will be supported by a reversed KM plot of the time to censoring, where the censoring indicator of the primary PFS analysis is reversed, to assess the number of subjects being followed over time.

4.2.7.2.5 Accounting for COVID-19 deaths

In the case that there is a sufficient number of subjects with a confirmed or suspected COVID-19 death event (either at least five subjects and/or at least 2% of the subject population), a sensitivity analysis will be conducted to assess the potential impact of COVID-19 deaths on PFS. This will be assessed by repeating the primary PFS analysis, but subjects who have not progressed prior to death and primary or secondary cause of death is due to COVID-19 infection, or a COVID-19 infection is reported as a fatal AE are censored at their last non-missing assessment prior to their COVID-19 infection death date.

4.2.7.3 Subgroup analyses of PFS

4.2.7.3.1 Subgroup analyses of PFS

The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors. This subgroup analysis will be conducted comparing PFS between osimertinib and placebo in the following subgroups of the FAS by BICR:

- Age group (<65 vs ≥65)
- Sex (Male vs Female)
- Smoking status (Current/Former (yes) vs Never (no))

- Prior chemoradiation strategy (CCRT vs SCRT)
- Disease stage prior to chemoradiation (IIIA vs IIIB/IIIC)
- Subjects enrolled at a Chinese site and declaring themselves of Chinese ethnicity vs subjects enrolled at non-Chinese site or subjects declaring themselves of non-Chinese ethnicity.
- Central plasma ctDNA EGFR (Ex19Del or L858R) mutation status at screening (positive vs negative vs unknown)
- Tissue EGFR mutation at screening (Ex19Del vs L858R)
- Race (Asian vs non-Asian)
- Response to prior CRT (CR vs PR vs SD vs NE)

The subgroup analyses for the stratification factors will be based on the values entered into the eCRF.

For each subgroup level of a factor, the HR and 95% CI will be calculated from a Cox proportional hazards model that only contains a term for treatment, the factor and treatment-by-factor interaction term. The Cox models will be fitted using PROC PHREG with the Efron method to control for 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.

If there are too few events available for a meaningful analysis of a particular subgroup comparison (it is not considered appropriate to present analyses where there are less than 20 events across both treatment groups in a subgroup comparison), the relationship between that subgroup and the primary endpoint (PFS) will not be formally analyzed. In this case, only descriptive summaries will be provided.

4.2.7.3.2 Quantitative interactions and consistency of treatment effect between subgroups

The presence of quantitative interactions will be assessed by means of an overall global interaction test.

This will be performed on the FAS by comparing the fit of a Cox proportional hazards model including treatment, covariates for prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC), and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity), and all covariate-by-treatment interaction terms, with one that excludes the interaction terms. The global interaction will be assessed at the 2-sided 10% significance level. If the fit of the model is not significantly improved, 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 & Simon 1985).

4.2.8 Secondary outcomes

The two endpoints of OS and CNS PFS will be tested after the primary PFS analysis in a hierarchical procedure at the time of the PFS analysis, following the primary analysis.

4.2.8.1 CNS progression-free survival (CNS PFS)

Time to CNS progression event, is the time from the randomization to the earliest event of CNS progression or death using CNS BICR assessments, according to RECIST v1.1. Subjects in the FAS will be analyzed as in the primary analysis of PFS using a log-rank test stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity) for generation of the p-value. The HR and 95% CI will be obtained as for the primary analysis of progression-free survival provided there are ≥20 CNS Progression events (CNS progression or death) across both treatment groups. A KM plot will also be presented by treatment group.

The cumulative incidence rate of CNS PFS, by BICR, at 12 and 24 months will be produced using the KM method with,

$$CI(t) = 1 - S(t)$$

where $CI(t)$ is the cumulative incidence rate at time t and $S(t)$ is the survival function at time t.

The above analyses will be repeated for the Investigator-assessed CNS PFS using RECIST v1.1 assessments as a sensitivity analysis.

In addition, a summary of disease characteristics at CNS progression (i.e. sites of new lesions) will be produced by treatment arm.

Competing risk analysis

The FAS will be used for this analysis. The analysis will need to combine data from CNS BICR and whole body BICR. Whole body BICR will contribute progressions outside the CNS whilst CNS BICR will contribute progressions in the CNS.

Once a subject experiences progression, no further scans are required per protocol. Therefore, subjects who have progression outside of the CNS, or die in the absence of progression prior to any potential progression in the CNS will not provide brain scans which may document a later CNS progression.

In these cases, a progression outside of the CNS or death in the absence of progression hinders the observation of the event of interest (CNS progression). In addition, by observing the non-CNS progression, the chance of a CNS progression may be altered.

To potentially account for this, competing events will also be considered in an analysis.

Three events will be defined:

1. Radiological documentation of progression in the CNS (event of interest). This will include subjects who had documented radiological progression in the CNS at the same overall visit.

2. Radiological documentation of progression outside of the CNS (competing event).

3. Death by any cause in the absence of the previous 2 events (competing event).

Subjects who experience progression in the CNS and in other anatomy at the same overall visit assessment could be defined as a fourth competing risk group. However, given the event of interest has been observed it is felt most appropriate to include these in the event of interest category.

Each subject will be assigned a value of 0 to 3, with 0 presenting subjects who haven't experienced any event yet (censored) and 1-3 indicating that an event has been observed as defined above (defined as the earliest if more than 1 does occur). For a subject who is censored in the analysis, the censoring time will be the latest of the dates contributing to a particular overall visit assessment. For subjects experiencing an event, the event time will be the first occurrence of the event.

The KM method will be used to produce the cumulative incidence function. Here the probability of any event happening is partitioned into the probabilities for each type of event where

$$\widehat{F_1}(t) + \widehat{F_2}(t) + \widehat{F_3}(t) = 1 - \hat{S}(t)$$

In addition, the 95% confidence interval for the event of interest ($\widehat{F_1}(t)$) will be calculated at 12- and 24-month time intervals. The confidence interval will be found for the log(-log

($\widehat{F_1}(t)$)) and then transformed back to the original scale to avoid the bounds being negative or greater than 1.

The cumulative incidence function will be produced for each treatment group, with 6 lines on one plot.

A summary will be provided to display the number of each type of event by treatment group. The treatment status at CNS progression of subjects at the time of analysis will be summarized. This will include the number (%) of subjects who were on treatment at the time of progression, the number (%) of subjects who discontinued study treatment prior to progression, the number (%) of subjects who have not progressed and were on treatment or discontinued treatment and the number of subjects whose CNS event was death.

In addition, a summary of subjects who progressed with and without CNS progression will be produced by treatment arm.

4.2.8.2 Overall survival (OS)

The analysis of OS will be conducted at two timepoints: at the time of the primary analysis of BICR-assessed PFS and at approximately 60% maturity when approximately 120 death events (across both arms) have occurred. Overall survival data will be analyzed using the same methodology and model as for the primary analysis of PFS provided there are sufficient events (≥20 deaths across both treatment groups with at least 5 events per arm) available for a meaningful analysis otherwise descriptive summaries will be provided. A KM plot will also be presented by treatment group.

OS rates will be reported with 95% confidence intervals at set time points e.g. 12 months, 24 months.

4.2.8.3 Sensitivity Analysis – Accounting for COVID-19 deaths

In the case that there is a sufficient number of subjects with a confirmed or suspected COVID-19 death event (either at least five subjects and/or at least 2% of the subject population), a sensitivity analysis will be conducted to assess the potential impact of COVID-19 deaths on OS. This will be assessed by repeating the primary OS analysis, but subjects who had primary or secondary cause of death is due to COVID-19 infection, or a COVID-19 infection is reported as a fatal AE, are censored at their last non-missing assessment prior to their COVID-19 infection death date.

4.2.8.4 Objective response rate (ORR)

Objective response rate by BICR will be analyzed using logistic regression stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity). All BICR scans will be used regardless of whether they were scheduled or not and both confirmed and unconfirmed responses will be used. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater

than 1 will favor osimertinib) together with its associated profile likelihood 95% CI (e.g. using the 'LRCI' option in GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). The ORR analysis will be repeated in the evaluable for response set, and for confirmed BICR ORR in FAS and in the evaluable for response set.

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

4.2.8.5 Duration of response (DoR)

Descriptive data will be provided for the DoR in subjects responding to treatment based on the BICR RECIST v1.1 assessments in the FAS, including the associated Kaplan-Meier curves and median DoR with associated 95% CI (without any formal comparison or p-value attached). All BICR scans will be used regardless of whether they were scheduled or not and both confirmed and unconfirmed responses will be used. The DoR analysis will be repeated for confirmed BICR.

4.2.8.6 Disease control rate (DCR)

Disease control rate is defined as the percentage of subjects who have a best overall response of CR, PR or SD as assessed by BICR.

Disease control rate, by BICR assessments, will be analyzed using logistic regression stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity). The results of the analysis will be presented in terms of an odds ratio together with its associated 95% profile likelihood CI and 2-sided p-value.

4.2.8.7 Depth of response

Depth of response (i.e. tumor shrinkage/change in tumor size) by BICR will be examined by summarizing the absolute change in target lesion tumor size from baseline, and percentage change in target lesion tumor size from baseline using descriptive statistics and presented at each time point and by randomized treatment group. Depth of response will also be examined by presenting the proportion of subjects who achieve >30%, >50% and >75% reduction in target tumor lesion size.

The best percentage change from baseline in target lesion tumor size will also be summarized descriptively and presented graphically using waterfall plots, with each subject's best percentage change in tumor 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 tumor size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. Bars will be color coded based on the unconfirmed BoR response category for each subject. Subjects with imputed

values based on the rules described in Section 3.2.2.7 will be marked with an asterisk. A different marker will indicate subjects where progression was due to a new lesion. A waterfall plot will also be produced by treatment arm, and by each stratification factor. Stratification factors will be based on the values entered into the eCRF.

The effect of osimertinib on best percentage change in tumor size will be estimated from an analysis of covariance (ANCOVA) model including a covariate for baseline TL tumor size, a covariate for the time from the baseline scan to randomization and a covariate for each of the stratification factors used in the primary PFS analysis.

The number of subjects, unadjusted mean, and least squares means for each treatment group will be presented, together with the treatment difference in least squares means, 95% CI and corresponding p-value.

4.2.8.8 Time to death or distant metastases and time to study treatment discontinuation (TTDM and TTD).

Time to death or distant metastases by BICR (TTDM) and time to study treatment discontinuation or death (TTD) for subjects in the FAS will be analyzed using a log-rank test stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity) for generation of the p-value. The HR and CI will be obtained according to the same method applied to the primary analysis of PFS.

Competing risk analysis

Competing risk analysis will also be performed for time to distant metastases or death (TTDM). The FAS will be used for this analysis.

In cases where a subject progresses or dies in the absence of progression, no further scans are required per protocol, and therefore subjects who progress locally, i.e. no new lesions detected outside the field of radiation, or who die in the absence of progression will not provide scans which may document distant metastases at a later point.

To potentially account for this, competing events will also be considered in an analysis where the three events will be defined as:

1. Distant metastases (event of interest). Defined as any new lesion that is detected on a scan (outside of the radiation field) according to RECIST v1.1 or proven by biopsy.

2. Radiological documentation of local progression (competing event). This is progression according to RECIST v1.1 but where no new lesions outside of radiation field identified.

3. Death by any cause in the absence of the previous 2 events (competing event).

The analysis will follow the method described in Section 0.

The KM method will be used to produce the cumulative incidence function and the 95% confidence interval for the event of interest will be calculated at 12- and 24-month time intervals.

The cumulative incidence function will be produced for each treatment group, with 6 lines on one plot.

A summary will be provided to display the number of each type of event by treatment group.

4.2.8.9 Post progression outcomes (PFS2, TFST and TSST)

Second progression-free survival on a subsequent treatment (PFS2), time to first subsequent treatment (TFST) and time to second subsequent treatment (TSST) for subjects in the FAS will be analyzed using a log rank test stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC) and China Cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity) for generation of the p-value. The HR and CI will be obtained according to the same method applied to the primary analysis of the PFS. KMs plot will also be presented by treatment group.

4.2.9 Patient reported outcomes

Endpoints based on EORTC QLC-C30 and EORTC QLC-LC13 scales are secondary outcomes; all other PRO endpoints are exploratory outcomes.

4.2.9.1 EORTC QLQ-C30 and EORTC QLQ-LC13

Compliance

Summary measures of compliance and evaluability over time will be derived for both the EORTC QLQ-C30 and EORTC QLQ-LC13 questionnaires. These will be based upon the received, expected and evaluable questionnaires and will be presented by visit and treatment arm.

Response by visit

Summary tables of visit responses (absolute scores, change from baseline, improved, worsened, and stable) for each of the symptoms, functional scales, and GHS/QoL scale in the EORTC QLQ-C30 and EORTC QLQ-LC13 for visits where >=20 subjects have data across both treatment groups, unless otherwise specified, will be presented by treatment group.

Time to symptom and HRQoL/functioning deterioration

For each of the symptom and functioning subscales and GHS/QoL subscale in the EORTC QLQ-C30 and the dyspnea, cough, and chest pain scores from the EORTC QLQ-LC13 time to deterioration will be analyzed.

This will be analyzed by performing a stratified log-rank test the same as that for the primary analysis of PFS stratified by prior chemoradiation strategy (CCRT vs SCRT), disease stage prior to chemoradiation (IIIA vs IIIB/IIIC), and China cohort (enrolled at a Chinese site and subject declaring themselves of Chinese ethnicity vs enrolled at non-Chinese site or subject declaring themselves of non-Chinese ethnicity). Summaries of the number and percentage of subjects who have an event as well as who were censored will be provided. KM medians of time to deterioration will be presented with 95% CIs for each treatment group along with percentage of subjects without deterioration at 6 months and 12 months and 95% CIs. The hazard ratio and it's corresponding 95% CI for each subscales will be presented.

Time to symptom deterioration will also be presented using a Kaplan-Meier plot for each of the symptoms, functioning subscales, and the GHS/QoL subscale from the EORTC QLQ-C30, and the dyspnea, cough, and chest pain scores from the EORTC QLQ-LC13.

Symptom improvement rate

The improvement rate for each of the symptom and functioning subscales, and the GHS/QoL subscale from EORTC QLQ-C30 and the dyspnea, cough, and chest pain scores from the EORTC QLQ-LC13 will be summarized. The symptom improvement rate for each score from each questionnaire will be compared by treatment group using logistic regression similar to the ORR analysis in Section 4.2.8.4.

Mixed models repeated measures of change from baseline

Change from baseline in each of the symptom, functioning, and GHS/QoL scores from the EORTC QLQ-C30 will be analyzed using an MMRM analysis of the post-baseline scores for visits with the use of data from baseline up to 10 months or date of PD (whichever is earlier), excluding visits with excessive missing data (defined as more than 75% missing data). Dyspnea, cough, and chest pain scores from the EORTC QLQ-LC13 will also be analyzed in the same way. The MMRM model will include subject, treatment, visit and treatment by visit interaction as explanatory variables, the baseline PRO score as a covariate along with the baseline PRO score by visit interaction. Treatment, visit and treatment by visit 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 visits 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 and p-value. The treatment-by-visit interaction will remain in the model regardless of significance.

For the purpose of the MMRM analysis, PRO visits will be windowed according to the principles set out in Section 4.1.1.

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.

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.

4.2.9.2 Subject Reported Outcome version of the Common Terminology Criteria for Adverse Event System (PRO-CTCAE)

Data from the PRO-CTCAE will be summarized on the SAF. The number of subjects with each level of response for each CTCAE item at visits where >=20 subjects have data, unless otherwise specified, will be summarized.

For symptoms that use branching logic, severity and interference responses will be excluded for subjects who reported "Never" for frequency. If subjects reported "None" for severity, then responses will not be summarized under the interference category if a response is entered.

4.2.9.3 Patients Global Impression of Severity (PGIS)

Response categories at each timepoint will be summarized descriptively on the FAS at each visit where >=20 subjects have data across both treatment groups.

4.2.9.4 EQ-5D-5L health state utility index

Descriptive statistics on the FAS will be reported for each health state domain (e.g. proportion in each domain) and the visual analogue scale by visit, as well as the change in the visual analogue scale value at each visit where >=20 subjects have data across both treatment groups.

4.2.9.5 Healthcare resource use module

The potential impact the disease and treatment have on health care resource use will be analyzed. Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) will be provided for each arm 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 subject presents with.

4.2.10 Safety

All safety data will be summarized using the SAF and listed only. No formal statistical analysis is planned for the safety data. For change from baseline summaries for vital signs,

laboratory data, LVEF, ECGs and physical examination, the baseline value will be the latest result obtained prior to the start of investigational product.

4.2.10.1 Laboratory assessments

All laboratory assessments will be listed. All values will be classified as low, normal or high based on the project-specific reference ranges and any values falling outside the reference ranges will be flagged as such.

For all continuous laboratory assessments, absolute value and percentage change from baseline will be summarized using descriptive statistics at each scheduled assessment time by actual treatment group. For categorical laboratory assessments, shift from baseline will be summarized using frequency and proportion at each scheduled assessment time by actual treatment group.

For clinical chemistry and hematology, shift tables of the number of CTCAE grade changes from baseline will be provided. Corresponding shift tables ("Negative", "Trace", "+", "++", "+++", "++++") will be produced for urinalysis.

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. A box plot of change from baseline for all hematology and clinical chemistry parameters will also be presented.

The denominator used in laboratory summaries will only include evaluable subjects, in other words those who had sufficient data to have the possibility of an abnormality. For example:

- If a CTCAE criterion involves a change from baseline, evaluable subjects would have both a pre-dose and at least 1 post-dose value recorded
- If a CTCAE criterion does not consider changes from baseline, to be evaluable the subject need only have 1 post dose-value recorded

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).

Hy's law (HL)

The following summaries will include the number (%) of subjects who have:

- Elevated ALT, AST, or Total bilirubin during the study
- ALT ≥3x ULN during the study.
- AST ≥3x ULN during the study.
- Total bilirubin ≥2x ULN during the study.
- ALT or AST ≥3x ULN during the study.

- ALT or AST ≥3x ULN and total bilirubin ≥2x ULN at any point during the study following the start of study treatment irrespective of an increase in ALP (potential Hy's law). The onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation.
- ALP ≥1.5x ULN
- ALP ≥3x ULN

Narratives will be provided in the CSR for subjects who have ALT ≥3x ULN plus total bilirubin ≥2x ULN or AST ≥3x ULN plus total bilirubin ≥ 2x ULN at any visit.

Liver biochemistry test results over time for subjects with elevated ALT (i.e. ≥3x ULN) or AST (i.e. ≥3x ULN), and elevated total bilirubin (i.e. ≥2x ULN) (at any time) will be plotted.

Individual subject data where ALT or AST plus total bilirubin are elevated at any time will be listed also.

Plots of maximum post-baseline ALT and AST vs. maximum post-baseline total bilirubin, expressed as multiples of ULN, will also be produced with reference lines at 3×ULN for ALT and AST, and 2×ULN for Total bilirubin. In each plot, total bilirubin will be in the vertical axis.

4.2.10.2 Physical examinations

Abnormalities identified from the physical examinations will be listed if present .

4.2.10.3 Vital signs

A box plot of change from baseline vital signs data are presented at each scheduled assessment time. All vital signs data will also be listed. The denominator in vital signs data should include only those subjects with recorded data.

4.2.10.4 Electrocardiograms (ECGs)

A box plot of change from baseline ECG data containing the ECG parameters R-R, PR, QRS, QT and QTcF will be presented at each scheduled assessment time. A listing of ECG test results for each parameter and visit will be produced.

4.2.10.5 Left ventricular ejection fraction (LVEF)

LVEF measurements will be summarized over time by treatment group and visit on the SAF by their absolute values and change from baseline. All LVEF results for the SAF will be listed.

4.2.10.6 WHO performance status

All WHO performance statuses will be summarized over time by treatment group on the SAF. A listing of the WHO performance status will be produced.

4.2.10.7 Adverse events (AEs)

All AEs in terms of the MedDRA preferred term and CTCAE grade will be listed and summarized by a count and percentage for each treatment group.

AEs occurring before start of treatment with study drug (i.e. before dose day 1) and which did not worsen during the study 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 AEs occurring after the end of the 28-day follow-up period or onset after the start of subsequent cancer therapy (including radiotherapy, except for palliative radiotherapy), or those occurring prior to treatment will be reported separately and not included in other AE summaries.

Summary information will be tabulated for:

- All AEs
- All AEs causally related (as assessed by the investigator) to study medication.
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related (as assessed by the investigator) to study medication.
- AEs with outcome of death
- AEs with outcome of death causally related (as assessed by the investigator) to study medication.
- All SAEs
- All SAEs for subjects receiving open-label osimertinib
- All SAEs causally related (as assessed by the investigator) to study medication.
- AEs leading to discontinuation of study medication.
- AEs leading to discontinuation of study medication, causally related (as assessed by the investigator) to study medication.
- AEs leading to dose modification (treatment interruption and/or dose reduction)

An overall summary of the number and percentage of subjects 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 subjects in any treatment group will be summarized by preferred term, by decreasing frequency. This cut-off may be modified after review of the data.

Other individual AE summaries of the number and percentage of subjects with the following adverse events will be produced. These will be summarized by system organ class, preferred term and actual treatment group:

- AE leading to dose reduction.
- AE leading to dose interruption.
- AE leading to discontinuation of study medication.
- Possibly related to IP.

AEs will be assigned CTCAE grades (National Cancer Institute (NCI) CTCAE version 5) and summaries of the number and percentage of subjects will be provided by maximum reported CTCAE grade, system organ class, preferred term and actual treatment group.

In addition, AEs with outcome of death, SAEs, AEs leading to discontinuation of treatment and AEs causally related to treatment will be listed.

AEs associated with COVID-19 and confirmed/suspected COVID-19 AEs are listed. Further summaries may be provided dependent on the numbers of such events.

The following listings will also be produced:

- Radiotherapy timing, for subjects with CTCAE grade 2 or higher pneumonitis, radiation pneumonitis or interstitial lung disease
- Radiotherapy timing for subjects with CTCAE grade 2 or higher cardiac events.

Deaths

A summary of deaths will be provided with the number and percentage of subjects by randomized treatment group. The total number of deaths, regardless of the date of death will be summarized. The deaths will be categorized according to the following mutually exclusive categories:

- Death related to disease under investigation **only** (determined by the investigator
- AE outcome of death only (not related to disease under investigation):
    - With AE onset date prior to initiation of subsequent anti-cancer therapy and <=28 days after last dose of study medication
    - With AE onset date after initiation of subsequent anti-cancer therapy or >28 days after last dose of study medication (whichever is earlier)
- AE outcome of death, related to disease under investigation:
    - With AE onset date prior to initiation of subsequent anti-cancer therapy and <=28 days after last dose of study medication

  - With AE onset date after initiation of subsequent anti-cancer therapy or >28 days after last dose of study medication (whichever is earlier)

- Death unrelated to AE or disease under investigation:
  - Death after initiation of subsequent anti-cancer therapy or >28 days after last dose of study medication [c]
- Subjects with unknown reason for death
- Other deaths [d]

If a subject has multiple fatal AEs, the earliest start date is used to determine the category they fall into, i.e. onset <=28 days after last dose of study medication or onset > 28 days after last dose of study medication.

Adverse Events of Special Interest

Preferred terms used to identify AESI will be listed before database lock (DBL) and documented in the Study 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.

The Patient Safety team will be responsible for maintaining the AESI list and up versioning when the MedDRA version changes.

Summary tables of the AESIs of cardiac failure, pneumonitis, interstitial lung disease and radiation pneumonitis will be produced. The number (%) of subjects experiencing any of the specified terms will be presented overall and by maximum CTCAE grade, and this summary will be further broken down by subjects within the Asia region, Non-Asia region, subjects randomized in Japan and subjects randomized outside of Japan within the Asia region.

Additional summaries of:

- time to onset of first AE for each grouped term and each preferred term within it
- time from last dose of radiotherapy to onset of first adverse event of special interest, for each grouped term and preferred terms
- time to onset of first CTCAE grade 3 or higher
- time to onset of first adverse event of special interest of CTCAE Grade 2 or higher and of Grade 3 or higher, for each grouped term and preferred term
- 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. A summary table of the number of subjects with any AESI whilst receiving open-label osimertinib will also be produced.

Summaries of the above-mentioned grouped AE categories will include the number and percentage of subjects who have:

- At least one AESI presented by outcome.
- At least one AESI casually related to study medication.
- At least one AESI leading to discontinuation of study medication.

A summary of the 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.

Prevalence plots by treatment arm, may be presented for each of the AESI grouped terms, provided there are ≥10 events, that is 10 events per each treatment arm. The prevalence at time t after first dose of study treatment is calculated as the number of subjects experiencing the event divided by the number of subjects receiving study treatment or in safety follow-up at time t; generally, t is categorized by each day after dosing. Multiple occurrences of the same event are considered for each subject, but a subject is only counted in the numerator whilst they are experiencing one of the occurrences of the event.

4.2.11 Pharmacokinetic data (PK)

The Pharmacokineticist will agree to the strategy for dealing with data affected by protocol deviations before any formal statistical analysis is performed. Important protocol deviations include changes to the procedures that may impact the quality of the data or any circumstances that can alter the evaluation of the PK. Examples include, but not limited to, vomiting following oral dosing occurring within the timeframe of 2 times the median time to reach maximum concentration (tmax); sample processing errors that lead to inaccurate bioanalytical results; incomplete dose administered; incomplete PK profile collected; and/or use of disallowed concomitant medication. In the case of an important protocol deviation or event, affected PK data collected will be excluded from the summaries and statistical analyses, but will still be reported in the study result listings. Important deviations will be listed and summarized in the CSR.

All plasma concentrations of osimertinib and AZ5104 will be listed.

The following summary statistics by nominal sample time will be presented on the PK analysis set for osimertinib and AZ5104:

- Geometric mean (calculated as $\exp(\mu)$, where μ is the mean of the data on a logarithmic scale)

- Coefficient of variance (CV, calculated as $\sqrt{[\exp(s^2) - 1]}$), where s is the standard deviation of the data on a logarithmic scale.
- Geometric mean ± standard deviation (calculated as $\exp(\mu) \pm s$)
- Arithmetic mean calculated using untransformed data.
- Standard deviation (sd) calculated using untransformed data.
- Minimum and maximum
- Number of observations (n)

PK concentration data from this study may be analyzed using a population PK approach and may also form part of a pooled analysis with other osimertinib studies; results from these analyses, if conducted may be reported separately from the CSR.

In addition, a listing of individual plasma/serum concentrations for osimertinib will be presented.

4.2.12 Other exploratory outcomes

4.2.12.1 Relationship between PK and other variables

Correlation of PK with other primary, secondary or exploratory endpoints in subjects treated with osimertinib will be examined. Results from such analyses maybe reported separately from the CSR, if warranted.

4.2.12.2 Comparison of baseline tumor EGFR mutation status

The following comparisons will be performed for baseline tumor EGFR mutation status if sufficient EGFRm data are available to allow meaningful exploratory analysis:

1. Tumor DNA vs plasma ctDNA (central **cobas®** result will be used as the reference where available, otherwise the local **cobas®** result will be used)

2. Local vs central **cobas®** EGFR mutation test (central **cobas®** result will be used as the reference)

Baseline tumor EGFR mutation status will be compared between tumor DNA and plasma ctDNA using the Kappa coefficient. Analysis will be carried out separately for each sensitizing mutation, Ex19Del and L858R, and as well as in aggregate. Analysis will be in all Part II screened subjects with evaluable results from baseline plasma samples. The central **cobas®** result will be used where available, otherwise the local **cobas®** result will be used as reference.

The baseline tumor EGFR mutation status will also be compared between the local and central **cobas®** EGFR mutation test v2, using the Kappa coefficient, in subjects with evaluable results from baseline tumor samples. Analysis will be carried out separately for each sensitizing mutation, Ex19Del and L858R, and as well as in aggregate.

The Kappa coefficient, κ will be calculated by

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the relative observed agreement among the two tests i.e. overall percent agreement (OPA), and p_e is the hypothetical probability of chance agreement (McHugh 2012):

$$p_e = \frac{\left(\frac{(TP + FN) \times (TP + FP)}{n}\right) + \left(\frac{(FP + TN) \times (FN + TN)}{n}\right)}{n}$$

Table 9 2x2 table format for reporting results comparing a new test outcome to the reference standard outcome

		Reference standard	
		+	–
New test	+	TP	FP
	–	FN	TN
Total		TP+FN	FP+TN
TP=number of True Positive events; FP=number of False Positive events; TN=number of True Negative events; FN=number of False Negative events.			

The following measures of agreement will also be calculated for both comparisons described above: The overall percent agreement (OPA); the percentage of total subjects where the new test and the reference standard agree. This will be calculated from Table 9 in the following way:

- Overall percent agreement (OPA) = 100% x (TP+TN)/(TP+FP+FN+TN)
- Positive percent agreement (PPA) = 100% x TP/(TP+FN)
- Negative percent agreement (NPA) = 100% x TN/(FP+TN)

The overall agreement will always lie somewhere between the positive percent agreement and the negative percent agreement.

The predictive value of a positive result (PPV) i.e. the proportion of test positive subjects who have the target condition and predictive value of a negative result (NPV) i.e. the

proportion of test negative subjects who do not have the target condition, will also be calculated.

- NPV= 100 x TN/(TN+FN)
- PPV= 100 x TP/(TP+FP)

Exact confidence intervals will be calculated and presented for each of the 5 measures for comparison of tissue vs plasma, but only for PPV for comparison of central vs local, using the Clopper-Pearson method to derive the 95% CIs.

5 INTERIM ANALYSES

5.1 Independent Data Monitoring Committee (IDMC)

An Independent Data Monitoring Committee (IDMC) will be convened, and will meet to review unblinded safety data, initially approximately 6 months after the study has started, as long as a minimum of 20 subjects have been randomized and treated for >1 month. Three subsequent meetings will take place every 6 months and then meetings will be held yearly thereafter until primary analysis completion. Further meetings for review of safety data from all subjects may be convened at the discretion of the IDMC. Following each meeting the IDMC will evaluate whether the trial should continue without change, be modified or stopped due to potential harm to subjects.

IDMC analysis will be presented on FAS and SAF where relevant.

5.2 Interim analysis for OS

One interim analysis for OS will be performed at the time of PFS analysis i.e. at approximately 60% maturity when approximately 120 PFS BICR-assessed events have occurred across both arms. See Section 4 for details on the MTP and alpha spending strategy.

6 REFERENCES

Aaronson et al 1993
Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365-76.

Amit, O. et al 2011
Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. European Journal of Cancer, 47(12), pp. 1772-8

Bergman et al 1994
Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use

in lung cancer clinical trials. EORTC Study Group on Quality of Life. Eur J Cancer 1994;30A:635-42.

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Statistics in Medicine 2009;28:739-61

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

EuroQol Group 2015
EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. URL https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf

Gail & Simon 1985
Gail M, Simon R. Testing for qualitative interactions between treatment effects and subject subsets. Biometrics. 1985; 41:361-72

McHugh ML 2012
McHugh ML. Interrater reliability: the kappa statistic. Biochem Med. 2012 Oct;22(3):276-282.

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

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

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: d5160c00048-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	Doc ID-003867659	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
19-Dec-2023 09:29 UTC	PPD	Content Approval
19-Dec-2023 15:45 UTC	PPD	Author Approval
19-Dec-2023 10:02 UTC	PPD	Content Approval

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