- **Protocol number:** D5160C00007.
- **Document title:** A Phase III, Double-Blind, Randomised Study to Assess the Efficacy and Safety of AZD9291 versus a Standard of Care Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor as First-Line Treatment in Patients with Epidermal Growth Factor Receptor Mutation Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer.
- **Version number:** 3.0.
- **Date of the document:** 28 February 2017.
- **NCT number:** NCT02296125.

AstraZeneca

Statistical Analysis Plan

Study Code	D5160C00007
Edition Number	3.0
Date	28 February 2017

A Phase III, Double-Blind, Randomised Study to Assess the Efficacy and Safety of AZD9291 versus a Standard of Care Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor as First-Line Treatment in Patients with Epidermal Growth Factor Receptor Mutation Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer

A Phase III, Double-Blind, Randomised Study to Assess the Efficacy and Safety of AZD9291 versus a Standard of Care Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor as First-Line Treatment in Patients with Epidermal Growth Factor Receptor Mutation Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer

Study Statistician PPD 02 Mar 2017
Date

A Phase III, Double-Blind, Randomised Study to Assess the Efficacy and Safety of AZD9291 versus a Standard of Care Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor as First-Line Treatment in Patients with Epidermal Growth Factor Receptor Mutation Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer

Global Product Statistician PPD 02/03/2017

Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE(s)	Adverse event(s)
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
B	Blood
BICR	Blinded Independent Central Review
BP	Blood pressure
BUN	Blood urea nitrogen
cEFR	CNS evaluable for response set
cFDA	China Health authority
CI(s)	Confidence interval(s)
CLIA	Clinical Laboratory Improvement Amendments
cMET	Proto-oncogene encoding Hepatocyte Growth Factor Receptor
CM	Concomitant medication
CNS	Central nervous system
CR	Complete response
CSR	Clinical Study Report
CT	Computer tomography
CTCAE	Common Terminology Criteria for Adverse Event
ctDNA	Circulating tumour deoxyribonucleic acid
CTSQ-16	Cancer Therapy Satisfaction Questionnaire - 16 items
CV	Coefficient of variation
DCR	Disease Control Rate
DNA	Deoxyribonucleic acid
DoR	Duration of Response
ECG	Electrocardiogram
eCRF	Electronic case report form

Abbreviation or special term	Explanation
EDoR	Expected Duration of Response
EDR	Early discrepancy rate
EGFR	Epidermal Growth Factor Receptor
EGFRm+	Epidermal Growth Factor Receptor mutation positive
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30 items
EORTC QLQ-LC13	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Lung Cancer 13 items
Ex19del	Deletions in exon 19
FAS	Full Analysis Set
FPI	First patient in
Gmean	Geometric mean
HbA1C	Haemoglobin A1C
HDU	High dependency unit
HER2	Human Epidermal Growth Factor Receptor 2
HP	Haybittle-Peto
HR	Hazard ratio
HRQoL	Health Related Quality of Life
ICU	Intensive care unit
IPCW	Inverse Probability of Censoring Weighting
KM	Kaplan-Meier
L858R	Exon 21 point mutation L858R
LDH	Lactate dehydrogenase
LDR	Late discrepancy rate
LLoQ	Lower limit of quantification
LOQ	Limit of quantification
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measures
MRI	Magnetic resonance imaging

Abbreviation or special term	Explanation
MTS	Metastases
NC	Not calculable
NE	Not evaluable
NPA	Negative percent agreement
NPV	Negative predictive value
NQ	Non-quantifiable
NSCLC	Non-small Cell Lung Cancer
NTL	Non-target lesion
OAE(s)	Other significant adverse event(s)
ORR	Objective Response Rate
OS	Overall Survival
P	Plasma
PD	Progressive disease
PFS	Progression free survival
PFS2	Second progression free survival; i.e. time from randomisation to second progression
PK	Pharmacokinetics
PPA	Positive percent agreement
PPV	Positive predictive value
PR	Partial response
PRO	Patient Reported Outcome
PRO-CTCAE	Patient Reported Outcome version of the Common Terminology Criteria for Adverse Event System
PT	Preferred term
QoL	Quality of Life
QT	Interval on the electrocardiogram representing the duration of depolarization and repolarization of the heart
QTc	The QT interval corrected for heart rate
QTcF	Fridericia QTc correction
RBC	Red blood cells

Abbreviation or special term	Explanation
RECIST	Response Evaluation Criteria in Solid Tumours
REML	Restricted maximum likelihood
RPSFT	Rank Preserving Structural Failure Time
S	Serum
SAE(s)	Serious adverse event(s)
SD	Stable disease
SMQ	Standardised MedDRA query
SoC	Standard of Care
SOC	System organ class
StdDev	Standard deviation
T790M	An amino acid substitution at position 790 in EGFR, from a threonine to a methionine
T790M+	T790M mutation positive
TFST	Time to first subsequent therapy
TKI	Tyrosine kinase inhibitor
TL	Target lesion
TSST	Time to second subsequent therapy or death
U	Urine
ULoQ	Upper limit of quantification
USA	United States of America
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief description of change
22 Feb 2017	Primary PRO measures are updated Multiple testing strategy updated The analysis method for HR and CI will be obtained directly from the U and V statistics The treatment effect HR and the two-sided 95% CIs will also be presented on the forest plot Added the Mixed models repeated measures of change from baseline in primary PRO symptoms Added the analysis method for tissue results based on method used for randomisation Updated the analysis methods section for the CTCAE Updated the Post progression, Best response on second treatment, Time on second treatment, Continuation of randomised treatment post investigator assessed RECIST progression, Time to treatment discontinuation, Time to discontinuation of any EGFR TKI sections Added the subgroup analysis comparing PFS between treatments by number of CNS lesions
16 Dec 2016	Added detail of additional prospective analysis of brain metastases (CNS). Multiple testing strategy updated to remove PFS subgroup in T790M+ patients and replace with a CNS PFS analysis. List of important protocol deviations for the CSR updated Method to determine HR and 95% CI for PFS and similar analyses updated to use Cox proportional hazards model instead of U and V statistics Duration of response updated to use log-normal instead of looking across 3 different models ctDNA section updated as only baseline ctDNA data will be presented in the CSR Removed clinically meaningful changes in EORTC QLQ-C30 and EORTC QLQ-LC13, and analyses of symptom improvement rate and time to symptom deterioration.

Date	Brief description of change
	Updated SAP in line with Global Protocol Amendments 1 and 2, as follows: • Updated BICR text to state that all patients' scans will undergo BICR. Removed algorithm for determination of the random sample, and added a summary of analyses repeated for the BICR; • Added details around possibility of crossover to AZD9291 for subjects randomised to SoC and details of analyses repeated for crossover subjects; • Added information for analyses of data from Japan; • Updated numbers for screened patients, randomised patients, progression events required for the PFS analysis and deaths required for the OS analysis, and other values used in the sample size calculations; • Updated tumour assessment timings; • Added details for analysis of overall survival adjusting for the impact of crossover subjects, to be completed if this treatment sequence occurs in a significant proportion of patients; • Clarified that unscheduled visits are included in derivation of maximum/minimum values; • Clarified process for calculating EORTC QLQ-LC13 dyspnoea scale score if not all values are recorded, per handbook; • Added compliance calculations and summaries for the EORTC QLQ-C30 and EORTC QLQ-LC13; • Clarified that no PK parameters will be calculated, as there are only two post-dose time points on each day; • Clarified that where randomisation stratification factors are included in an analysis model, the variables used at randomisation (and not the values recorded in the eCRF) will be used; • Clarified that baseline for PROs is the last observation prior to the first dose of study treatment; • Updated multiple testing diagram and text for clarity; • Added EGFR-TKI comparator selection to baseline characteristics; • Clarified the PFS analysis time points as 6, 12, 18 and 24 months; • Updated attrition bias sensitivity analysis text to match protocol; • Clarified the OS analysis time points as 6, 12, 18 and 24 months for the primary PFS analysis, and 12, 24 and 36 months for the survival follow up; • Removed the summary of patients with abnormalities at baseline and post-baseline for each physical examination body system.
09 Feb 2015	N/A

1. STUDY DETAILS

1.1 Study objectives

Primary Objective:	Outcome Measure:
To assess the efficacy of single agent AZD9291 compared with Standard of Care (SoC) Epidermal Growth Factor Receptor (EGFR)-tyrosine kinase inhibitor (TKI) therapy as measured by progression free survival (PFS).	PFS according to Response Evaluation Criteria in Solid Tumours (RECIST) v1.1 by Investigator assessment.

Secondary Objectives:	Outcome Measure:
To assess the efficacy of AZD9291 compared with SoC EGFR-TKI therapy by assessment of PFS in patients with: - Positive (or negative) pre-treatment T790M (amino acid substitution at position 790 in EGFR, from a threonine to a methionine) mutation. - EGFR Ex19del or Exon 21 (L858R) mutation. - Epidermal Growth Factor Receptor mutation positive (EGFRm+) Deletions in exon 19 (Ex19del) or L858R detectable in plasma-derived circulating tumour deoxyribonucleic acid (ctDNA).	PFS according to RECIST v1.1 by Investigator assessment.
To further assess the efficacy of AZD9291 compared with SoC EGFR-TKI therapy.	- Objective Response Rate (ORR) - Duration of Response (DoR) - Disease Control Rate (DCR) - Depth of response All according to RECIST v1.1 using Investigator assessments.
To further assess the efficacy of AZD9291 compared with SoC EGFR-TKI therapy.	Overall survival (OS)
To characterise the pharmacokinetics (PK) of AZD9291 and its metabolites (AZ5104 and AZ7550).	Plasma concentrations of AZD9291 and metabolites AZ5104 and AZ7550; and ratio of metabolite to AZD9291 at predose and 0.5 to 2 hours and 3 to 5 hours postdose.

Secondary Objectives:	Outcome Measure:
To assess the impact of AZD9291 compared to SoC EGFR-TKI therapy on patients' disease-related symptoms and Health Related Quality of Life (HRQoL).	- Change from baseline in European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30 items (EORTC QLQ-C30) - Change from baseline in European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Lung Cancer 13 items (EORTC QLQ-LC13). Please see Appendix E of the study protocol for details of the questionnaires and versions.
To assess patient satisfaction with treatment when receiving AZD9291 compared with SoC EGFR-TKI therapy.	Cancer Therapy Satisfaction Questionnaire 16 items (CTSQ-16) Please see Appendix E of the study protocol for details of the questionnaire and version.

Safety Objective:	Outcome Measure:
To assess the safety and tolerability profile of AZD9291 compared with SoC EGFR-TKI therapy.	- Adverse events (AEs) graded by Common Terminology Criteria for Adverse Event (CTCAE) version 4.0. - Clinical chemistry, haematology, and urinalysis. - Vital signs, physical examination, body weight. - Digital electrocardiogram (ECG). - Left Ventricular Ejection Fraction (LVEF). - World Health Organization (WHO) Performance Status. - Ophthalmologic assessment.

Exploratory Objective*:	Outcome Measure:
CCI	

* Results from such exploratory analyses may be reported separately from the CSR.

Exploratory Objective*:	Outcome Measure:
CCI	
T CCI	
	CCI
CCI	- CCI
CCI	CCI
CCI	CCI
CCI	CCI

† Data generated may be reported separately and may also form part of a pooled analysis with other AZD9291 studies.
‡ Samples may be analysed retrospectively. Any biomarker data generated may be reported separately and may also form part of a pooled analysis with other AZD9291 studies.

Exploratory Objective*:	Outcome Measure:
CCI	CCI

1.2 Study design

This is a Phase III, double-blind, randomised study to assess the efficacy and safety of AZD9291 (80 mg orally, once daily) versus SoC, EGFR-TKI (either gefitinib [250 mg orally, once daily] or erlotinib [150 mg orally, once daily]), as first-line treatment in patients with locally or centrally confirmed EGFRm+, locally advanced or metastatic Non-small Cell Lung Cancer (NSCLC), not amenable to curative surgery or radiotherapy.

Patients will be enrolled based on either a locally available EGFR mutation result, which has been performed in a Clinical Laboratory Improvement Amendments (CLIA)-certified (United States of America [USA] sites) or an accredited laboratory (outside of the USA) or by testing performed at a designated central laboratory. All patients who are enrolled based on locally available EGFR mutation results or who are tested centrally for enrolment, will be required to provide biopsy tissue for central testing of the two most common EGFR mutations known to be associated with EGFR-TKI sensitivity (Ex19del and L858R substitution mutation). This is to allow a sensitivity analysis to be performed for the various local testing methods used to recruit patients to the study by comparing local testing results with the central laboratory. The EGFR mutation status of the patient's tumour will be determined by the designated central laboratory using the cobas® EGFR Mutation Test (Roche).

Patients should continue with their randomised treatment until RECIST v1.1 defined progression or until a treatment discontinuation criterion is met. There is no maximum duration of treatment as patients may continue to receive their randomised treatment beyond RECIST v1.1 defined progression as long as they are continuing to show clinical benefit, as judged by the Investigator. Patients will be followed for progression and survival.

The primary endpoint for this study is PFS (defined by RECIST v1.1), as assessed by the Investigator. Progression free survival has been chosen as a clinically meaningful outcome measure, representing a direct benefit to the patient that is largely unaffected by the effects of subsequent therapy. The sponsor will be assessing OS as a key secondary endpoint recognising that OS is an important objective assessment of clinical benefit, however in this treatment naive population, OS will likely be confounded by the use of subsequent therapies including protocol permitted cross-over.

The agreement between the treatment effect assessed by the investigators and the blinded independent central review will be assessed by all patients' scans having a Blinded Independent Central Review (BICR).

On discontinuation of randomised study drug, patients will be treated in accordance with the regional SoC.

Following objective disease progression according to RECIST 1.1, as per investigator assessment, patients who were randomised to the control arm may have the option to receive open-label AZD9291, provided the following criteria are met and the patient wishes to do so:

- Disease progression confirmed by independent central imaging review which must be established prior to a patient being unblinded. (Note: if progression is not confirmed centrally, the patient is not eligible to receive open-label AZD9291 at that time. Should it be in the patient's best interests, they may continue to receive randomised treatment and submit the next scan for central imaging review according to the schedule.)

- No intervening therapy following discontinuation of randomised treatment.

- Tumour confirmed as T790M mutation positive (T790M+) following disease progression. (This may be determined before or after a patient has been unblinded.)

Provided the above criteria have been met, and the patient was randomized to the control arm, the patient may commence open-label AZD9291. Patients who have been unblinded prior to central confirmation of progression are not able to receive open-label AZD9291. If patients are not eligible for crossover or choose not to crossover, they cannot recommence or continue on their randomised treatment.

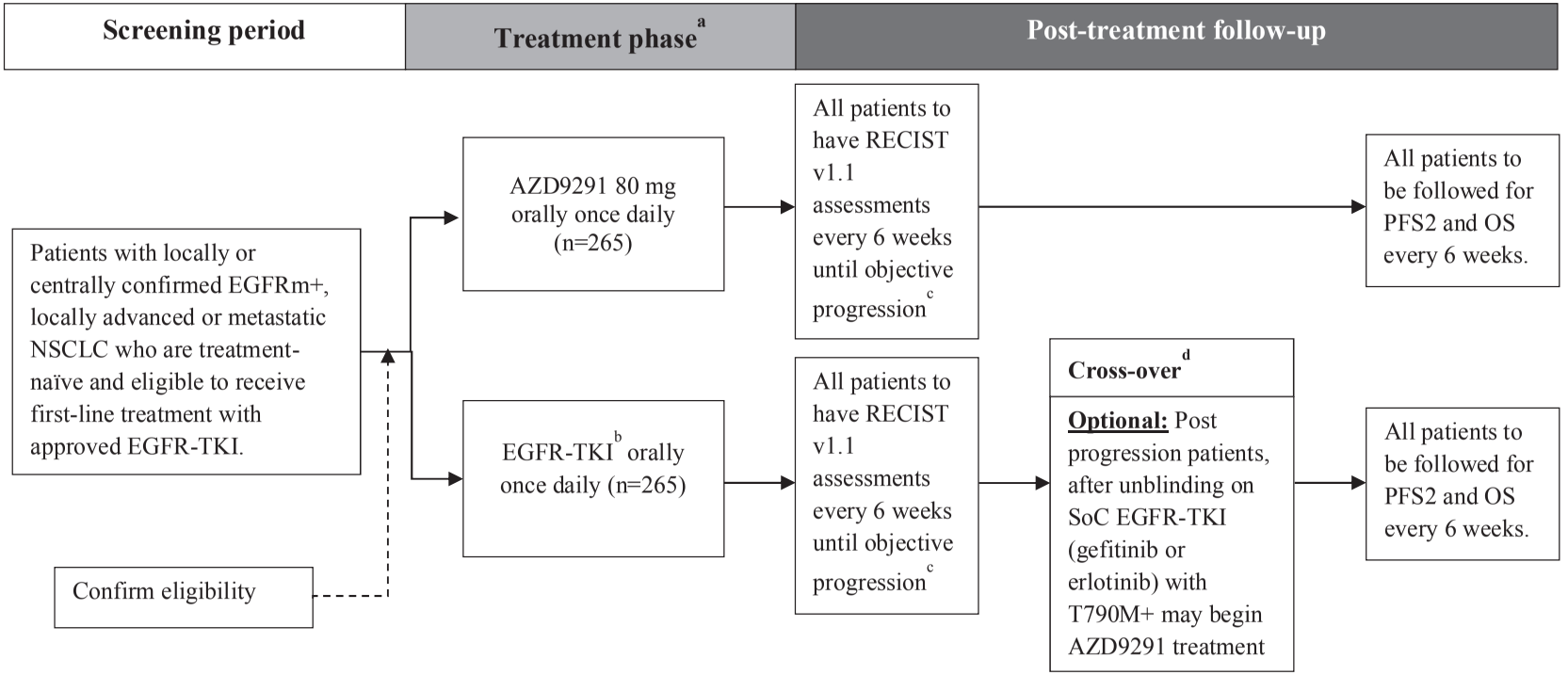
After the data cut-off date for the primary PFS analysis, all patients (except those enrolled in China) determined to have objective disease progression according to RECIST 1.1, as per investigator assessment, will be given the opportunity to begin treatment with open-label AZD9291, if eligible. Patient's tumour must have been confirmed as T790M mutation positive from biologic material collected post progression; central confirmation of disease progression will no longer be required. All patients enrolled in China will be given the opportunity to begin treatment with open-label AZD9291, if eligible, without the requirement for central confirmation of disease progression, after the data cut-off date for the China PFS analysis.

It is expected that data from 120 Chinese patients will be required to support a submission in China. Some patients (estimated to be approximately 30 patients) will be recruited in China as part of the 530 globally randomised patients and will be included in the analysis of the global study. The remaining patients (estimated to be approximately 90 patients) will be recruited after enrolment of the global study has ended and will not be included in the analysis of the global study. These additional patients will only be enrolled from mainland China. A separate analysis of all patients recruited in China (i.e. patients recruited in China from either the Global study or the China Extension) will be performed and is described in Appendix 1.

Additionally, a separate analysis of all patients recruited in Japan will be performed at the time of the global PFS analysis.

The overall study design is shown in Figure 1 below. The study schedule is detailed in Section 4 of the protocol.

Figure 1 Study flow chart



a Patients will continue to receive study drug until objective disease progression or as long as they are continuing to show clinical benefit, as judged by the investigator.

b Either gefitinib (250 mg orally, once daily) or erlotinib (150 mg orally, once daily).

c Patients who discontinue treatment prior to disease progression will continue to have RECIST v1.1 assessment every six weeks for the first 18 months and then every 12 weeks until objective progression. Patients who continue treatment after objective progression due to clinical benefit will be followed up as per standard practice post progression.

d Patients with objective radiological progression according to RECIST 1.1 by the Investigator and confirmed by independent central imaging review who are on SoC EGFR-TKI (gefitinib or erlotinib) after being unblinded and have T790M+ will be given the opportunity to cross-over and begin treatment with AZD9291 80mg, once daily. After data cut-off date for the primary PFS analysis, all patients (except those enrolled in China) determined to have objective

disease progression according to RECIST 1.1 as per Investigator's assessment will be given the opportunity to begin treatment with open-label AZD9291, if eligible; central confirmation of disease progression will no longer be required. The patients enrolled in China will be given the opportunity to begin open-label treatment with AZD9291 (if eligible) after the primary PFS analysis for the China patient subgroup.

1.3 Number of patients

Approximately 530 patients will be randomised globally in a 1:1 ratio (AZD9291: SoC EGFR TKI) to this study. The primary endpoint of the study is PFS based on Investigator assessment (according to RECIST v1.1). Progression free survival analysis will be performed at approximately 29 months after first patient in (FPI) for 12 months recruitment (or 30 months for 15 months recruitment).

The primary analysis of PFS will occur when approximately 359 progression events have been observed in the 530 globally randomised patients. If the true PFS hazard ratio (HR) for the comparison of AZD9291 versus SoC EGFR TKI is 0.71, 359 progression events will provide 90% power to demonstrate a statistically significant difference in PFS at a 5% two-sided significance level (translating to an approximate improvement in median PFS from 10 to 14.1 months assuming exponential data distribution and proportional hazards). The minimum critical HR is 0.81 (i.e. 10 to 12 months).

In order to randomise 530 patients, it is estimated that 980 EGFRm+ patients will need to be screened.

For the key secondary endpoint of PFS in patients with T790M+ NSCLC confirmed using a highly sensitive assay, there will be approximately 72% power to detect a PFS HR=0.55 (i.e. 10 to 18 months), assuming a prevalence of 20%.

For the OS analysis, there will be approximately 72% power to demonstrate a HR <0.75 (i.e. 25 to 33.3 months) with two-sided 5% significance level. If the 4 month improvement as detailed for PFS is maintained for OS, then there will be approximately 27% power to demonstrate a HR of 0.86 (median OS of 25 months for SOC and 29 months for AZD9291) with two-sided 5% significance level.

Following randomisation of 530 patients in the global portion of the study, additional mainland China patients (until approximately 120 in total) will be randomized to facilitate the China-only analysis dataset.

2. ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 Full analysis set

The Full analysis set (FAS) will include all globally randomised patients. Any patients recruited in China only, after global recruitment has ended, will not be included in the FAS (see Appendix 1). Patients who were randomised but did not subsequently receive treatment are included in the FAS. Global recruitment is defined as ending on the last date that a patient outside mainland China completes screening (i.e. is randomised or screen failed).

The FAS will be used for all efficacy analyses and treatment groups will be compared on the basis of randomised study treatment, regardless of the treatment actually received.

2.1.2 Safety analysis set

The safety analysis set will consist of all globally recruited patients who received at least one dose of study treatment. Any patients recruited in China only, after global recruitment has ended, will not be included in the safety analysis set (see Appendix 1).

Safety data will not be formally analysed but summarised using the safety analysis set, according to the treatment received; i.e. erroneously treated patients (e.g. those randomised to treatment A but actually given treatment B) will be summarised according to the treatment they actually received. If a patient receives both treatments then they will be summarised according to the treatment they were randomised to.

2.1.3 Pharmacokinetic analysis set

Pharmacokinetic Analysis Set is defined as patients in the FAS who have at least one evaluable PK concentration (see below) and who have no detectable pre-dose AZD9291 concentrations above the LLQ on Cycle 1 Day 1.

An evaluable PK concentration is supported by the relevant date and time of this sample. For Cycle 1, Day 1, a PK sample obtained with the corresponding dosing data for that day; and for samples taken after multiple dosing (Cycle 2 and beyond), the dosing data for the 2 days prior to the sample day as well as the sample day must be available. For multiple dosing PK concentrations, the patients should have taken the protocol defined dose of 80 mg daily for at least 7 consecutive days before the PK sample was obtained without any dose interruption or reduction. For any individual sample from a patient to be included in the PK analysis, the full sample data and dosing data need to be present for that sample/patient.

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 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 summarised in the clinical study report (CSR).

2.1.4 Centrally confirmed EGFRm analysis set [tissue]

The centrally confirmed EGFRm Analysis Set is defined as patients in the FAS who have centrally confirmed EGFRm+ with either Ex19del or L858R substitution mutations. Patients who were screened/randomised under a local test and sent a tissue sample for the central test will be included according to the central test result.

2.2 Violations and deviations

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

A list of important protocol deviations and any action to be taken regarding the exclusion of patients or affected data from specific analyses are defined in Table 1 below and a list of deviations that are regarded as important are in Table 2 below.

Note that the contents of these tables are not an exhaustive list. A complete list of anticipated protocol deviations (including important protocol deviations) will be compiled separately and finalised prior to unblinding.

Table 1 Protocol deviations with action to be taken for analysis

Protocol Deviation	Action to be Taken For Analysis
Patient did not receive any study medication	Exclude from the Safety analysis set
Patient was given incorrect study medication	Analyse "As-randomised" for the FAS. Analyse "As-treated" for the Safety analysis set. If only the incorrect study medication was taken for the whole study, then analyse according to treatment taken. If patient received both AZD9291 and SOC then analyse according to randomised treatment.

Table 2 Important deviations

Criteria type	Important Deviation Description
Inclusion	No pathologically confirmed adenocarcinoma of the lung
Inclusion	No locally advanced or metastatic NSCLC, not amenable to curative surgery or radiotherapy.
Inclusion	No confirmation that the tumour harbours an EGFR mutation known to be associated with EGFR TKI sensitivity (including exon 19 deletion and L858R).
Inclusion	No lesions, not previously irradiated and not chosen for biopsy during the study screening period, that can be accurately measured at baseline as ≥ 10mm in the longest diameter (except lymph nodes which must have short axis ≥ 15mm) with computerised tomography (CT) or magnetic resonance imaging (MRI) which is suitable for accurate repeated measurements.
Inclusion	Patient is not treatment-naïve for locally advanced or metastatic NSCLC.

Criteria type	Important Deviation Description
Exclusion	a) Prior treatment with any systemic anti-cancer therapy for advanced NSCLC including standard chemotherapy, biologic therapy, immunotherapy, or any investigational drug. b) Prior treatment with an EGFR-TKI. c) Alternative anti-cancer treatment. d) Treatment with an investigational drug within five half-lives of the compound or any of its related material, if known.
Exclusion	Past medical history of ILD, drug-induced ILD, radiation pneumonitis which required steroid treatment, or any evidence of clinically active ILD.
IP Administration/ Study treatment	Patient received / used incorrect investigational product
IP Administration/ Study treatment	Patient restarted study treatment after experiencing corneal ulceration or Interstitial Lung Disease (ILD)
IP Administration/ Study treatment	Patient did not receive any study medication but was randomised
Disallowed medications	Other anticancer agents, investigational agents or radiotherapy (for reason other than bone metastases) which are prohibited while the patient is on study treatment
Procedures / tests	Baseline tumour assessments (RECIST1.1) performed more than 28 days before randomization.
Procedures / tests	RECIST scans performed outside of the scheduled window on more than 2 occasions
Procedures / tests	Methods/Procedures for tumour assessment are not compliant with CSP or RECIST1.1.
Procedures / tests	Missing RECIST assessments for efficacy

The following summary will be provided:

- A summary of the number and percentage of patients with an important protocol deviation by treatment group and overall and by type of deviation (Analysis set: FAS)

A by-subject listing of important protocol deviations will be provided.

3 PRIMARY AND SECONDARY VARIABLES

3.1 General variables

3.1.1 Study day definitions

For the purpose of efficacy data summary, Study Day 1 is defined as the date of randomisation to study treatment. For visits (or events) that occur on or after randomisation, study day is defined as (date of visit [event] - date of randomisation + 1). For visits (or events) that occur

prior to randomisation, study day is defined as (date of visit [event] - date of randomisation). There is no Study Day 0.

For the purpose of safety data summary, Dose Day 1 is defined as the date of first dose of study treatment (referred to in the protocol as Week 1 Day 1). For visits (or events) that occur on or after first dose, dose day is defined as (date of visit [event] - date of first dose of study treatment + 1). For visits (or events) that occur prior to first dose, dose day is defined as (date of visit [event] - date of first dose of study treatment). There is no Dose Day 0.

For listings (such as for AEs) that include the derivation of "days since last dose," this is defined as (event date - date of last dose). Events that occur on the same day as the last dose of study drug will therefore be described as occurring zero days from the last dose of study drug.

3.1.2 Visit windows

For summaries of vital signs, laboratory data, ECG, HRQoL, and PROs etc., assessments will be assigned to calculated visit windows (using study day).

The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the visit window should be based on the actual date and not the intended date of the visit. 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.

The window for the visits following baseline (including unscheduled visits) will be constructed in such a way that the upper limit of the interval falls half way between the two visits.

For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval). Values from scheduled and unscheduled visits will be included. Listings should display all values contributing to a time point for a subject; they should also highlight the value for that subject that was used in the summary table, wherever feasible.

For visit-based summaries:

- If there is more than one value per subject within a visit window then the closest to the planned study day value should be summarised, or the earlier in the event the values are equidistant from the planned study day. The visit will be missing if no assessment was reported within the specified visit window around the planned study day.
- To prevent very large tables or plots being produced that contain many cells with meaningless data, summary statistics will be presented where at least 10 patients in either treatment group have data recorded at a particular visit.

3.1.3 Handling missing data

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

3.1.3.1 Imputation of partial dates

Concomitant medication and adverse events start dates

- If year is missing (or completely missing), do not impute.
- If (year is present and month and day are missing) or (year and day are present and month is missing), impute as January 1st.
- If year and month are present and day is missing, impute day as first day of the month.

Concomitant medication and adverse events end dates

- If year is missing (or completely missing), do not impute.
- If (year is present and month and day are missing) or (year and day are present and month is missing, impute as December 31st, unless this is after the date of death in which case date of death will be used instead.
- If year and month are present and day is missing, impute day as last day of the month, unless this is after the date of death in which case date of death will be used instead.

In addition for AEs and CMs if, for a partial start date, the start date could (when also considering the end date) potentially be on the first study medication date, the start date will be imputed with the first study medication date to assume a "worst case" scenario; e.g. AE from UNK-Feb-2014 to 23-Mar-2014 with first study medication date 21-Feb-2014, then the AE start date will be imputed to 21-Feb-2014.

3.1.4 Imputation rules for lab values outside of quantification range

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

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

3.1.5 Patient reported outcome version of the common terminology criteria for adverse event system grouped terms (PRO-CTCAE)

The grouped terms for PRO-CTCAE are:

- Cutaneous:
  - Rash, Skin Dryness, Acne, Hair Loss, Hand-Foot Syndrome, Itching, Nail Loss, Nail Ridging, Nail Discoloration

- Cardio/Circulatory:
  - Swelling
- Gastro-Intestinal:
  - Taste Changes, Decreased Appetite, Nausea, Vomiting, Constipation, Diarrhoea, Abdominal Pain, Fecal Incontinence
- Sleep/Wake (1 item):
  - Fatigue
- Neurological:
  - Numbness and Tingling
- Visual/Perceptual:
  - Blurred Vision
- Oral:
  - Dry mouth, Mouth/Throat Sores, Cracking at the corners of the mouth (Cheliosis)
- Others:
  - Sensitivity to sunlight, Nose bleed, Bruising, and Chills

3.2 Efficacy assessments

3.2.1 Investigator RECIST-based assessments

From the investigators review of the imaging scans, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST 1.1. It will also be used to determine the endpoints ORR, DoR, DCR and PFS from the overall visit response and scan dates. Responses of CR and PR do not require confirmation in line with RECIST v1.1 criteria for randomised trials. The endpoint of depth of response will be assessed from tumour size measurements based on investigator assessment.

Baseline radiological tumour assessments are to be performed no more than 28 days before the start of randomisation. Tumour assessments are then performed every six weeks following randomisation for the first 18 months and then every 12 weeks until disease progression. Baseline values recorded after randomisation should not be used as the baseline assessment; although such assessments can be used in the calculation of progressive disease (PD).

At each visit, patients will be programmatically assigned an overall visit response of complete response (CR), partial response (PR), stable disease (SD) or PD depending on the status of their disease compared with baseline and previous assessments. If a patient has had a tumour assessment, which cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE) (unless there is evidence of progression in which case the response will be assigned as PD).

Please refer to Appendix 2 for the definitions of CR, PR, SD and PD and the derivations of overall visit response using the information from target lesions (TL), non-target lesions (NTL) and new lesions.

Per the AZ guideline for RECIST v1.1, all central nervous system (CNS) lesions present at baseline are considered as non-target lesions.

3.2.1.1 Rounding of target lesion data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to 1 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%.

3.2.1.2 Missing TL data

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

- A new lesion is recorded;
- A NTL visit response of PD is recorded;
- The sum of TLs is sufficiently increased to result in a 20% increase, and an absolute increase of ≥5mm from nadir even assuming the non-recorded TLs have disappeared.

The nadir (i.e. smallest measurement) can only be taken from assessments where all the TLs had a longest diameter (LD) recorded.

3.2.1.3 Lymph nodes

For lymph nodes, if the size reduces to <10mm, 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 over-written as a CR.

3.2.1.4 TL visit responses subsequent to CR

A CR response can only be followed by CR, PD or NE. If a CR has occurred, 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), the response will be set to CR irrespective of whether the criteria for PD of TL is also met for lymph node (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 <10mmfor lymph nodes), the response will be set to NE irrespective of whether when referencing the sum of TL diameters the criteria for PD is also met.

- Step 3: If not all lesions meet the CR criteria, and the sum of lesions meets the criteria for PD, the response will be set to PD. (See Appendix 2 for the definitions of CR, PR, SD and PD).

- Step 4: If, after steps 1 through 3, a response cannot be determined, the response will be set to remain as CR.

3.2.1.5 TL too big to measure

If a TL becomes too big to measure, this should be indicated in the database and an estimated 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.

3.2.1.6 TL too small to measure

If a TL becomes too small to measure a value of 5mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and entered a smaller value that can be measured reliably. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

3.2.1.7 Irradiated lesions/lesion intervention

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

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation/palliative surgery/embolization), should be handled in the following way:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.
- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and scale up (based on the sizes at the nadir visit to give an estimated sum of diameters) as long as there remain ≤1/3 of the TLs with missing measurements. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then a scaled sum of diameters will be calculated, treating the lesion with intervention as missing, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention also has a value of 0 recorded.

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).

Once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours.

3.2.1.8 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.

3.2.1.9 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.

3.2.1.10 Change in method of assessment of TLs

Computerized tomography scan (CT) and magnetic resonance imaging (MRI) are the only methods of assessment that can be used within this trial. 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.

3.2.2 Blinded independent central review of RECIST-based assessments

The BICR of radiological imaging data will be carried out using RECIST v1.1. All radiological scans for all patients (including those at unscheduled visits, or outside visit windows) will be provided to the BICR. Up to three qualified radiologists will independently review all imaging scans in the following way. First a primary review will be performed by two independent radiologists for each patient, on a time point by time point basis to give an overall tumour assessment at each time point using RECIST 1.1. Then a global radiology review will be performed whereby the same independent radiologists will globally assess all time points for a patient in the review period and adjust an overall assessment if necessary. Finally, if the overall assessment for at least one time point for a patient does not agree between the two independent radiologists, a third independent radiologist will adjudicate and identify which radiologist's assessments they agree with and should be used (for the whole patient). The independent reviewers will be blinded to treatment.

Tumour assessment will be performed using contrast enhanced CT or MRI of chest and abdomen (including liver and adrenal glands) and other regions as clinically indicated. Duplicate images will be collected for the BICR. For each patient, the BICR will define the

overall visit response data (CR, PR, SD, PD, or NE) and the relevant scan dates for each time point. If a patient has had a tumour assessment which cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is any evidence of progression in which case the response will be assigned as PD).

Per the AZ guideline for RECIST v1.1, all CNS lesions present at baseline are considered as non-target lesions.

Further details of the BICR will be documented in the Independent Review Charter.

3.2.3 Progression free survival

Progression-free survival is defined as the time from randomisation until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to progression. Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST assessment.

However, if the patient progresses or dies after two or more missed visits, the patient will be censored at the time of the latest evaluable RECIST assessment.

Given the scheduled visit assessment scheme (i.e. six-weekly for the first 78 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 497 (i.e. week 71) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to twelve-weekly (i.e. between days 497 and 553) this will equate to 20 weeks (i.e. take the average of 6 and 12 weeks which gives 9 weeks and then apply same rationale, hence 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks). From week 79 onwards (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 patient has no evaluable visits or does not have baseline data, they will be censored at 1 days unless they die within two visits of baseline (on or before study day 91 (2 x 6 weeks x 7 +7 days for late assessment)).

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:

- Date of progression will be determined based on the earliest of the dates of the component that triggered the progression.

- When censoring a patient for PFS, the patient will be censored at the latest of the dates contributing to a particular overall visit assessment.

3.2.4 Objective response rate

Objective Response rate is defined as the number (%) of randomised patients with at least one visit response of CR or PR. Data obtained up until progression or last evaluable assessment in the absence of progression will be included in the assessment of ORR. However, any CR or PR which occurred after a further anti-cancer therapy (excluding radiotherapy) was received will not be included in the numerator for the ORR calculation (where the FAS will be the denominator). Responses of CR and PR do not require confirmation in line with RECIST v1.1 criteria for randomised trials.

3.2.5 Duration of response

Duration of response will be defined as the time from the date of first documented response until the date of documented progression or death in the absence of disease progression. The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of PR or CR.

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

3.2.6 Disease control rate

Disease control rate is defined as the percentage of patients who have a best overall response of CR or PR or SD at ≥6 weeks, prior to any PD event. The 6 week time point will allow for a visit window and be defined as on or after study day 35 (allowing for the visit window).

The best overall response for each patient will be determined as defined in Appendix 2.

3.2.7 Depth of response

Depth of response is defined as the relative change in the sum of the longest diameters of RECIST target lesions at the nadir in the absence of new lesions or progression of non-target lesions compared to baseline. The absolute change and percentage change from baseline in the sum of tumour size at each assessment will be calculated. The best change in tumour 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 patient has no TLs at baseline), a value of +20% will be imputed as the best percentage change from baseline in the following situations (otherwise best percentage change will be left as missing):

- If a subject has no post-baseline assessment 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

3.2.8 Overall survival

OS is defined as the time from the date of randomisation until death due to any cause. Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive.

The date the patient was last known to be alive will be taken from the survival status CRF form. The date of contact will not be used. Other CRF pages will not be considered.

Note: Survival calls will be made in the 2 weeks following the date of the data cut-off for each OS analysis, and if subjects are confirmed to be alive, or if the death date is post the final data cut-off date, these subjects will be censored at the date of the final data cut off. Death dates may be found by checking publicly available death registries.

3.2.9 Time from randomisation to second progression on subsequent treatment (PFS2)

Time from randomisation to second progression (PFS2) is defined as the time from the date of randomisation to the earliest of the progression event subsequent to that used for the primary variable PFS or date of death after starting subsequent anti-cancer treatment. If a subject dies without any progression events, the patient's PFS and PFS2 event dates would be equivalent. If a patient dies after their primary PFS event, but prior to the initiation of subsequent anti-cancer therapy their death date is still considered their PFS2 event. Patients 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/disease assessment date (or day 1 if no post-baseline RECIST data is available) if the patient has not had a second progression or death.

3.2.10 Time to first subsequent therapy or death

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

3.2.11 Time to second subsequent therapy or death

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

3.2.12 Brain metastases

The number of patients reporting brain metastases during the study treatment will be summarised based on investigator data where new lesions in the brain have been identified.

3.2.13 Symptoms and HRQoL

PROs will be assessed using the EORTC QLQ-C30, EORTC QLQ-LC13, CTSQ-16 and PRO-CTCAE questionnaires.

3.2.13.1 EORTC QLQ-C30 and EORTC QLQ-LC13

The EORTC QLQ-C30 consists of 30 questions, which can be combined to produce five functional scales (Physical, Role, Cognitive, Emotional, Social), three symptom scales (Fatigue, Pain, Nausea/vomiting), six individual items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties) and a global measure of health status/QoL. The EORTC QLQ-LC13 is a lung cancer specific module comprising 13 questions to assess lung cancer symptoms (cough, haemoptysis, dyspnoea, Pain in Chest, Pain in Arm or Shoulder and Pain in other Parts), treatment related side-effects (sore mouth, dysphagia, peripheral neuropathy and alopecia) and pain medication. With the exception of a multi-item scale for dyspnoea, all are single items.

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

Higher scores on the global health status and functioning scales indicate better health status/function. Higher scores on the symptoms scales indicate greater symptom burden. Note that the global health status scale is based on only the two specific HRQoL items and not the entire questionnaire.

For all scales other than dyspnoea scale in the EORTC QLQ-LC13, if at least half the components of a scale are present for a time point then the score will be calculated, otherwise the score will be set to missing. For the EORTC QLQ-LC13 dyspnoea scale, the score will only be calculated if all three items have been answered, otherwise the dyspnoea score will be set to missing and the items will be used as single-item measures.

The primary PRO measures will be five patient-reported lung cancer symptoms assessed using the EORTC QLQ-LC13 and C30, namely:

- Dyspnoea: (multi-item scale based on three questions: “Were you short of breath when you rested; walked; climbed stairs”), LC13
- Cough: (“How much did you cough?”), LC13
- Pain: (“Have you had pain in your chest?”), LC13

- Appetite Loss: ("Have you lacked appetite?") C30
- Fatigue: (multi-item scale based on three questions: "Have you felt weak?", "Did you need to rest?" "Were you tired?") C30

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:

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

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

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

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

3.2.13.2 Cancer therapy satisfaction questionnaire

The CTSQ-16 is a 16 item questionnaire measuring three domains related to patients' satisfaction with cancer therapy: Expectations of Therapy, Feelings about Side Effects and Satisfaction with Therapy.

3.2.13.3 Patient reported outcome common terminology criteria for adverse event symptoms

The PRO-CTCAE questionnaire will be used to derive subject reporting of treatment possibly related symptoms.

The PRO-CTCAE will only be administered in those countries where a linguistically validated version exists (languages available are English, German, Japanese and Spanish).
PRO-CTCAE is an item-bank of symptoms of the CTCAE experienced by subjects while undergoing treatment of their cancer that has been converted to subject terms (i.e. CTCAE term "myalgia" converted to "aching muscles"). Twenty-eight items from the PRO-CTCAE item-bank have been selected as being relevant for this study.

3.2.14 Health resource utilisation

Health resource utilisation will be assessed in terms of hospitalisation, outpatient visits and emergency department visits.

3.3 Safety variables

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

3.3.1 Adverse events

Adverse events (both in terms of Medical Dictionary for Regulatory Activities [MedDRA] preferred terms and CTCAE grade) will be listed individually by patient.

Any AE occurring before study treatment will be included in the data listings but will not be included in the summary tables of AEs.

Any AE occurring within 28 days of discontinuation of IP (i.e. the last dose of study treatment) will be included in the AE summaries. Any events in this period that occur after a patient has received further therapy for cancer (following discontinuation of study treatment) will be flagged in the data listings.

Preferred terms used to identify adverse events of special interest will be listed before database lock and documented in the Study Master File. Groupings of certain MedDRA preferred terms will be based on preferred terms provided by the medical team and a listing of the preferred terms in each grouping will be produced. The grouped terms expected are as follows: skin effects, nail effects, ILD and pneumonitis, ocular effects, diarrhoea, cardiac effects (QT), cardiac effects (C.F.) and upper GI tract inflammatory events.

3.3.2 Serious adverse events

Serious adverse events (SAEs) (both in terms of MedDRA preferred terms and CTCAE grade) will be listed individually by patient.

A SAE is an AE occurring during any study phase (i.e. screening, treatment, washout, follow-up), that fulfils one or more of the following criteria:

- Results in death.
- Is immediately life-threatening.
- Requires in-patient hospitalisation or prolongation of existing hospitalisation.
- Results in persistent or significant disability/incapacity or substantial disruption of the ability to conduct normal life functions.
- Is a congenital abnormality or birth defect.

- Is an important medical event that may jeopardise the patient or may require medical intervention to prevent one of the outcomes listed above.

Any SAE occurring before study treatment will be included in the data listings but will not be included in the summary tables of AEs.

Any SAE occurring within 28 days of discontinuation of IP (i.e. the last dose of study randomised or crossover treatment) will be included in the relevant SAE summaries. Any events in this period that occur after a patient has received further therapy for cancer (following discontinuation of study treatment) will be flagged in the data listings.

3.3.3 Other significant adverse events (OAEs)

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and AEs leading to discontinuation. Based on the expert's judgment, significant AEs of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered OAEs and reported as such in the CSR. A similar review of laboratory/vital signs (pulse and BP)/ECG data will be performed for identification of OAEs.

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

3.3.4 Duration of exposure

Total exposure time (months) will be calculated from the first dose to the last dose:

$$\text{Total exposure} = \frac{(last\ dose\ date\ where\ dose\ > 0mg\ -\ first\ dose\ date)\ +\ 1}{30.4375}$$

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

$$\text{Actual exposure} = \frac{\left((last\ dose\ date\ where\ dose\ > 0mg\ -\ first\ dose\ date)\ +\ 1\right) - total\ duration\ of\ dose\ interruption\ (i.e.number\ of\ days\ with\ dose\ =\ 0mg)}{30.4375}$$

3.3.5 Laboratory safety variables

The following laboratory variables will be summarised:

Table 3 Laboratory safety variables

Clinical chemistry (Serum [S] / Plasma [P])	Haematology (Blood [B])	Urinalysis (Urine [U])
S/P-Albumin	B-Haemoglobin	U-Glucose

Clinical chemistry (Serum [S] / Plasma [P])	Haematology (Blood [B])	Urinalysis (Urine [U])
S/P-ALT	B-Leukocyte	U-Protein
S/P-AST	B-Haematocrit	U-Blood
S/P-Alkaline phosphatase	B-Red blood cells (RBC) count	
S/P-Bilirubin, total	B-Absolute leukocyte differential count:	
S/P-Calcium, total	Neutrophils	
S/P-Corrected calcium	Lymphocytes	
S/P-Creatinine	Monocytes	
Creatinine Clearance[§]	Basophils	
S/P-Glucose (fasting, on PK days only)[**]	Eosinophils	
S/P-LDH[††]	B-Platelet count	
S/P- Haemoglobin A1C (HbA1C)	B-Reticulocytes	
S/P-Magnesium		
S/P-Potassium		
S/P-Sodium		
S/P-Urea nitrogen/Blood urea nitrogen (BUN)		

$$Corrected\ calcium = serum\ calcium + 0.8 \times (4 - serum\ albumin)$$

3.3.6 ECG

Fridericia QTc correction (QTcF) will be calculated as:

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

3.4 Pharmacokinetics variables

The plasma concentration data for AZD9291 and its metabolites AZ5104 and AZ7550 will be performed by Covance laboratories on behalf of Quantitative Clinical Pharmacology, AstraZeneca.

The ratio of individual metabolite to AZD9291 will be calculated at each time point. No PK parameters will be calculated as there are only two post-dose time points on each day.

[§] Creatinine clearance will be derived using the method of Cockcroft and Gault (Cockcroft & Gault, 1976).
[**] Patients will be required to fast (water only) for at least eight hours prior to the collection of a fasting glucose sample required on PK days. Random glucose sample will be collected on non-PK days.
[††] LDH is an additional variable collected at Visit 1 only.

The data collected in this study may also be combined with data from other studies and analysed using population PK and/or PK-pharmacodynamic methods. A separate analysis plan will be written to describe these analyses. The results of any such analyses will be reported separately from the CSR.

4 ANALYSIS METHODS

4.1 Planned analyses

The following table (Table 4) summarises the planned summaries and analyses.

Table 4 Planned summaries and analyses

Demographics and patient characteristics	
	Demography and baseline characteristics
	Medical history
	Prior and concomitant treatments
	Baseline Brain metastases
Efficacy, Quality of Life (QoL) and PK	
	PFS[‡‡§§]
	ORR[‡‡§§], DoR[‡‡§§], BOR[‡‡§§], DCR[‡‡], depth of response[‡‡§§]
	OS
	EORTC-QLQ-C30/LC13
	CTSQ-16
	PRO-CTCAE
	PK
	PFS2, TFST, TSST
Safety	
	Exposure
	Adverse events
	Serious adverse events
	Adverse events of special interest
	Deaths
	Laboratory evaluations
	Vital signs
	Physical examination
	ECG
	LVEF
	WHO performance status

[‡‡] Based on investigator assessment using RECIST 1.1.
[§§] Repeated using RECIST 1.1 by BICR assessment.

4.2 General principles

Continuous data will be summarised using descriptive statistics (number of observations, mean, standard deviation, median, 25th and 75th percentiles, minimum, and maximum). Frequencies and percentages will be used for summarising categorical (discrete) data.

Confidence intervals (CIs), when presented, will generally be constructed at the 95% level. For binomial variables, the normal approximation methods will be employed unless otherwise specified.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days.

Data will be presented in data listings by subject identifier and treatment arm.

All summaries will be presented by treatment group unless otherwise specified.

Where analysis models are stratified by the randomisation stratification factors race and mutation status (or the stratification factors are included in the model as covariates), the strata obtained at randomisation will be used, not the values recorded in the electronic case report form (eCRF).

Additional summaries of efficacy and other variables may be produced as a separate report(s) for specific regions (e.g. Japan), as required by local health authorities.

4.2.1 Baseline measurements and change from baseline variables

In general, for efficacy endpoints the last observed measurement prior to randomisation will be considered the baseline measurement. For safety and PRO endpoints the last observation prior to the first dose of study treatment will be considered the baseline measurement unless otherwise specified.

For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

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

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. For % change from baseline, calculate as:

$$\frac{(\text{post-baseline value} - \text{baseline value})}{\text{baseline value}} \times 100$$

4.2.2 Multiple testing strategy

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

The analyses of PFS, OS and CNS PFS endpoints will occur at the time of the primary analysis of PFS. 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 significance testing of CNS PFS in will be not be performed.

One analysis of the primary endpoint (PFS) is planned. Two analyses of OS are planned; one interim at the time of PFS and a final analysis. The final OS analysis is planned to be conducted when the OS data is approximately 60% mature (approximately 318 deaths).

A 2-sided 5% alpha will be used in all testing, with the exception of overall survival endpoint. Since two analyses of OS are planned, the Lan DeMets approach that approximates the O'Brien and Fleming spending function will be used to 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 SoC arm of 25 months and a median OS of 29 months on the AZD9291 arm, that 140 OS events were observed at the first analysis, the information fraction would be entered as 0.44 (140/318 events) for the first analysis since 140 events are expected at the interim analysis. This would result in a significance level for the first analysis of 0.001 (2-sided) and a significance level for the final analysis of 0.0496 (2-sided).

The information fraction is calculated as the number of events at the analysis time-point divided by the total number of events at the final analysis time-point.

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

4.3 Analysis methods

4.3.1 Subject disposition and data sets analysed

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

- Enrolled (informed consent received)

- Randomised
- Treated
- Patients ongoing study treatment at the data cut-off
- Patients randomised to receive SoC who crossed over to AZD9291
- Included in each analysis set (FAS, Safety, and PK).

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

A separate summary of disposition of crossover subjects, after crossing over to AZD9291, will be provided.

4.3.2 Protocol deviations

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

4.3.3 Demographic and other baseline characteristics

Demographic and baseline subject characteristics will be listed and summarised 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)

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

- Age group (years) (grouped as <50, ≥50-<65, ≥65-<75, ≥75)
- Sex
- Race
- Ethnic group

- Extent of Disease Upon Entry to Study (metastatic, locally advanced, both)
- Site of local/metastatic disease
- Brain metastases and Visceral metastases
- Baseline TL size (mean and categories: <40, 40-79, 80-119, ≥120mm)
- Histology type
- Smoking status
- WHO performance status (0/1)
- EGFR mutation type from cobas central testing
- EGFR mutation type from test for randomisation strata
- EGFR-TKI comparator selection (gefitinib, erlotinib)

A listing of all subjects showing all the EGFR mutations identified by the cobas® central test will be presented. The EGFR mutations identified by the local test will also be presented where applicable.

The number of patients recruited in each country and each centre will be presented by treatment group and in total.

4.3.4 Medical history

Disease related medical history and relevant surgical history will be coded using MedDRA. All disease related medical history (past and current) will be listed and the number and percentage of patients with any disease related medical history will be summarised for the FAS by system organ class (SOC) and preferred term (PT).

All relevant surgical history will be listed and summarised similarly.

4.3.5 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 eCRF.

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

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

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates will be imputed as detailed in Section 3.1.3.1.

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

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

In addition, all post-treatment anti-cancer medications and surgical procedures will be summarised for the full analysis set.

The following summaries will be produced:

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

All concomitant and other treatment data will be listed.

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

4.3.6 Exposure

Exposure will be summarised for the safety analysis set. The following summaries will be produced (and repeated for crossover patients):

- Summary of duration of exposure of study treatment
- Summary of interruptions and reductions of study treatment

4.3.7 Brain metastases

The proportion of patients who report brain metastases (either as sole new lesion or at the same time as progression on non-CNS lesions) during the trial will be summarised by treatment group.

The proportion of RECIST progression events by medical history of brain metastases will be summarized.

4.3.8 Efficacy

All efficacy analyses will be performed on the FAS. Results of all statistical analyses will be presented using a 95% CI and two-sided p-value.

4.3.8.1 Primary outcome: progression free survival

Progression free survival established from programmatically calculated response (see Section 3.2) based upon the Investigator assessment for patients in the FAS will be analysed using a log rank test stratified by race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R) for generation of the p-value and using the Breslow approach for handling ties. The HR and CI will be obtained directly from the U and V statistics as follows (Berry, et al., 1991), (Selke & Siegnumd, 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 standard deviation obtained from the LIFETEST procedure with a STRATA term for the stratification variable.

The covariates in the statistical modelling will be based on the values entered into IVRS at randomisation, even if it is subsequently discovered that these values were incorrect.

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

total number of events, median PFS (calculated from the Kaplan-Meier plot, with 95% CIs), and the percentage PFS at 6, 12, 18 and 24 months will be summarised. The progression status at the time of the PFS analysis will also be summarised, including the number and percentage of patients who progressed due to RECIST progression or due to death, or did not progress due to being progression free or due to being lost to follow up.

Sensitivity analyses

a) Ascertainment bias

The possibility of bias in assessment and measurement by Investigators will be assessed using the BICR assessment of disease progression by RECIST. Ascertainment bias will be assessed through the use of two measures proposed by Amit (Amit, et al., 2011): the early discrepancy rate (EDR) and late discrepancy rate (LDR). The EDR represents the positive predictive value of Investigator assessment and quantifies the frequency with which the Investigators declare progression early relative to BICR within each arm as a proportion of the total number of Investigator assessed PD's. The LDR quantifies the frequency that the Investigators declare progression later than BICR as a proportion of the total number of discrepancies within the arm. If the distribution of discrepancies is similar between the arms, then this suggests the absence of evaluation bias favouring a particular arm.

EDR and LDR are calculated as:

$$EDR = \frac{\begin{pmatrix}\text{number of times Inv. declares PD when BICR does not } + \\ \text{number of times Inv. declares PD earlier than BICR}\end{pmatrix}}{\begin{pmatrix}\text{number of times both Inv. and BICR declare PD} \\ + \text{ number of times Inv.declares PD when BICR does not}\end{pmatrix}}$$

$$LDR = \frac{\begin{pmatrix}\text{number of times BICR declares PD when Inv. does not } + \\ \text{number of times Inv. declares PD later than BICR}\end{pmatrix}}{\begin{pmatrix}\text{number of times Inv. declares PD when BICR does not} \\ + \text{ number of times BICR declares PD when Inv. does not} \\ + \text{ number of times Inv. declares PD later than BICR} \\ + \text{ number of times Inv. declares PD earlier than BICR}\end{pmatrix}}$$

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 investigators favouring the experimental arm.

BICR of all patients will be completed before database lock for the primary analysis of PFS.

b) 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 evaluable RECIST assessment will be analysed using a log rank test stratified by race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R), as described for the primary analysis of PFS.

For patients who die in the absence of progression, the date of death will be used to derive the PFS time used in the analysis. Patients with no events (i.e. patients censored for PFS) will be censored at day 1 for this sensitivity analysis.

c) Attrition bias

(i) Attrition bias will be assessed by repeating the primary PFS analysis, except that the actual PFS event times rather than the censored times of patients who progressed or died in the absence of progression immediately following two or more non-evaluable tumour assessments, will be included. In addition, and within the same analysis, patients who take subsequent therapy prior to their last evaluable RECIST assessment or progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy.

(ii) This analysis will be supported by a Kaplan-Meier plot of the time to censoring using the PFS data from the primary analysis, where the censoring indicator of the primary PFS analysis is reversed, will be presented to assess the number of patients being followed over time.

Subgroup analysis

In addition to the analysis of PFS described above, the following subgroup analyses will be conducted by comparing PFS between treatments (using a Cox-Proportional Hazards Model) in the following groups:

- Gender (Male / Female)
- Race (Asian / Non-Asian)
- Age at screening (<65 / ≥65)
- History of or current Brain metastases at entry (yes/no)
- Smoking history
- Baseline WHO Performance Status

- Pre-treatment T790M status (positive / negative)[***]
- EGFR mutation (Ex19del / L858R)[***]
- EGFR+ by ctDNA[***]
- Centrally confirmed EGFR+[***]

For each subgroup, the HR and 95% CI will be calculated from a single Cox proportional hazards model that contains a term for treatment, the subgroup covariate of interest and the treatment by subgroup interaction term. The treatment effect HR will be obtained for each level of the subgroup from this model. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties.

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

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

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

Quantitative interactions

The presence of quantitative interactions will be assessed by means of an overall global interaction test. This will be performed in the overall population by comparing the fit of a Cox proportional hazards model including treatment, covariates for race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R), and all covariate by treatment interaction terms, with one that excludes the interaction terms and will be assessed at the two-sided 10% significance level. If the fit of the model is not significantly improved (i.e. not statistically significant), then it will be concluded that overall, the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and any newly significant interactions re-included until a final model is reached where all included interactions are significant and all excluded interactions are non-significant. Throughout this process, all main effects will be included in

[***] Further details of these analyses are given in subsequent sections.

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).

[redacted]

4.3.8.2 Secondary outcomes

Hierarchical testing of key secondary variables

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 (see Section 4.2.2). The CNS PFS analysis is described in Section 4.3.9.

Analysis of PFS subgroups

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

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

PFS by T790M status subgroup [tissue]

Progression free survival in patients by confirmed pre-treatment T790M status (positive/negative) will be analysed using a Cox Proportional Hazards Model including treatment, T790M status, and the treatment by T790M status interaction term. The results of the analysis will be presented in terms of a HR together with its associated 95% CI and 2-sided p-value for positive patients and separately for negative patients.

PFS by EGFR Ex19del or L858R subgroup [tissue]

Progression free survival in patients by EGFR Ex19del or L858R substitution mutations (prospectively stratified) will be analysed using a Cox Proportional Hazards Model including treatment, mutation type (Ex19del / L858R), and the treatment by mutation type interaction term. The results of the analysis will be presented in terms of a HR together with its associated 95% CI and 2-sided p-value for patients with Ex19del and separately for patients with L858R.

Patients will be classified as EGFR Ex19del or L858R based on the IVRS data.

PFS by ctDNA subgroup [plasma]

An analysis of PFS including all patients by EGFR+ (either Ex19del or L858R positive mutations), EGFR- or invalid or no sample, identified from ctDNA will also be conducted using a Cox-Proportional Hazards Model including treatment, ctDNA type, and the treatment by ctDNA type interaction term. The results of the analysis will be presented in terms of a HR together with its associated 95% CI and 2-sided p-value for each level of ctDNA type.

PFS by Baseline brain metastases subgroup

Progression free survival in patients by baseline brain metastases will be analysed using a Cox Proportional Hazards Model including treatment, baseline brain metastases (Yes / No), and the treatment by baseline brain metastases interaction term. The results of the analysis will be presented in terms of a HR together with its associated 95% CI and 2-sided p-value for patients with baseline brain metastases and separately for patients without baseline brain metastases.

Analysis of overall survival

The analysis of OS will be conducted at two time points:

- At the time of the primary analysis of PFS.
- At approximately 60% maturity when approximately 318 death events (across both arms) have occurred in the FAS. It is predicted that 318 death events will be observed at approximately 45 months from FPI for 12 months recruitment (47 months for 15 months recruitment). Updated analyses of data collected post progression and safety may also be performed at this time.

The alpha will be split between the two analyses to provide strong control of the family-wise error rate (see Section 4.2.2).

Overall survival data will be analysed using the same methodology and model as for the analysis of PFS provided there are sufficient events available for a meaningful analysis (>20 deaths [if not, descriptive summaries will be provided]). The percentage OS at 6, 12, 18 and 24 months will be summarised for the analysis of OS at the primary PFS analysis, and the percentage OS at 12, 24 and 36 months will be summarised for the analysis of OS at the final survival follow up.

Additional analysis of overall survival adjusting for the impact of patients randomised to the control arm who crossover and subsequently receive AZD9291 will be completed if this treatment sequence occurs in a significant proportion of patients. Methods such as Rank Preserving Structural Failure Time (RPSFT) (Robins & Tsiatis, 1991), Inverse Probability of Censoring Weighting (IPCW) (Robins, 1993) and other methods in development may be explored. The decision to adjust and final choice of methods will be based on the plausibility

of the underlying assumptions. This additional analysis may be reported separately from the CSR.

Analysis of overall response rate

Overall response rate will be analysed using a logistic regression stratified by race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R). The results of the analysis will be presented in terms of an odds ratio together with its associated 95% profile likelihood CI and two-sided p-value.

The p-value will be based on a test statistics that is calculated as twice the change in log-likelihood resulting from the addition of a treatment factor to a model that contains the covariates defined above.

A summary of all responses will also be presented by treatment group.

Overall response rate will also be summarised by each stratification factor.

Analysis of duration of response

In order to analyse the secondary outcome variable of DoR between arms, the Expected Duration of Response (EDoR) will be derived for each treatment arm (Ellis, et al., 2008). The EDoR is the product of the proportion of patients responding to treatment and the mean DoR in responding patients, and provides an estimate based on all randomised patients.

Treatments will be compared by calculating the ratio of EDoRs using the log Normal probability distribution for duration of response in responding subjects. The log Normal distribution has been chosen based on a review of the fit of this distribution, the Exponential distribution, the Weibull distribution and the log logistic distribution to duration of response data from pooled data from the two Phase II AZD9291 single-arm monotherapy studies: D5160C0001C (AURA extension) and D5160C00002 (AURA2), using the data cut-off date of 1st November 2015. The log Normal distribution had the lowest value Akaike information criterion and Bayesian Information Criterion and the fitted line for the model overlaid with the KM plot indicated visually that the distribution was a reasonable fit. The mean duration of response will be estimated, and the null hypothesis that the distribution of the ratio of EDoRs is 1 will be tested and a p-value and 95% CI provided (Ellis, et al., 2008).

The analysis of DoR will be stratified by the same covariates as the primary analysis, weighting each stratum inversely proportional to the within stratum variance of the log of the ratio of EDoRs. Additionally, descriptive data will be provided for the DoR in responding patients, including associated Kaplan-Meier curves (without any formal comparison or p-value attached).

Analysis of disease control rate

Disease control rate will be analysed using a logistic regression stratified by race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R). The results of the analysis

will be presented in terms of an odds ratio together with its associated 95% profile likelihood CI and two-sided p-value.

The p-value will be based on twice the change in log-likelihood resulting from the addition of a treatment factor to a model that contains the covariates defined above.

Analysis of depth of response

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

The best percentage change from baseline in target lesion tumour size will also be summarised descriptively and presented graphically using waterfall plots, with each patient's best percentage change in tumour size represented as a separate bar and the bars ordered from the largest increase to the largest decrease. Reference lines at the +20% and –30% change in tumour size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. A waterfall plot will also be produced by each stratification factor.

The effect of AZD9291 on best percentage change in tumour size will be estimated from an analysis of covariance (ANCOVA) model with covariates for race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R), baseline tumour size and time from baseline scan to randomisation. The number of patients, unadjusted mean, and least squares means for each treatment group will be presented, together with the difference in least squares means, 95% CI and corresponding p-value.

Patient reported outcomes (PROs)

<u>EORTC QLQ-C30 and EORTC QLQ-LC13 Compliance Rates</u>

Summary measures of overall compliance and compliance over time will be derived for both EORTC QLQ-C30 and EORTC QLQ-LC13. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.
- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not died, is not lost to follow-up or withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under PRO follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.
- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing.

- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all time points, divided by total number of questionnaires expected to be received across all time points multiplied by 100.
- Overall patient compliance rate is defined for each randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.

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

Mixed models repeated measures of change from baseline in primary PRO symptoms

The analysis population for PRO data will be the FAS. Change from baseline in the primary PRO symptom scores of dyspnoea (EORTC QLQ-LC13), cough (EORTC QLQ-LC13), pain in chest (EORTC QLQ-LC13), fatigue (EORTC QLQ-C30) and appetite loss EORTC QLQ-LC30) will be regarded as the primary analysis of the PRO questionnaire data and will be analysed using a mixed model for repeated measures (MMRM) analysis of the change from baseline (defined in section 4.2.1) in PRO score for each visit.

The primary analysis will be to compare the average treatment effect from the point of randomisation for the first nine months (which will include visit data obtained at protocol scheduled time-points of baseline, days 8, 15, 22, 43, 64-106, 127-274 and the discontinuation and follow-up visits if occurring within the first nine months) unless there is excessive missing data (defined as >75% missing data in either arm).

It is acknowledged that patients will discontinue treatment at different timepoints during the study and that this is an important time with regards to PRO data collection. To account for this and in order to include the discontinuation and follow up visits, a visit variable will be derived for each subject in order that the average treatment effect can be analysed using the above method. Each visit will be assigned a sequential number. The time from randomisation to each of these will be derived in order to select only those visits occurring within the first nine months (274+7 days allowing for a window = 281 days) of randomisation.

As an example, say a patient (X) attends the first five scheduled visits and then discontinues treatment, whilst another patient (Y) discontinues treatment after the first two scheduled visits, the first eight generic visits would be as follows:

Generic visit	Days since randomisation	
	Patient X	Patient Y
Baseline	Baseline	Baseline
1	8	7

Generic visit	Days since randomisation	
2	15	16
3	22	21 (discontinuation)
4	29	30 (safety follow-up)
5	36	
6	43	
7	64 (safety follow-up)	
8	85(discontinuation)	

For Patients X and Y the first eight and four visits respectively could be used in the analysis of deriving the average treatment effect (change from baseline in PRO score) as they are within the first nine months of randomisation. Time windows as outlined in section 3.1.2 will be exhaustive so that data recorded at any time point has the potential to be summarised and included in the model. If there are two or more values potentially allocated to the same visit, the post baseline assessment closest to the scheduled visit date will be included in the summaries and in the MMRM.

The MMRM model will include patient, treatment, visit (generic) 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; patient 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. No p-values will be presented. The treatment-by-visit interaction will remain in the model regardless of significance.

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

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

where TRT is the randomised treatment, VISIT is the visit, PROSC is the change from baseline in the PRO score, and PROBL is the baseline PRO score.

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

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

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

Descriptive statistics and graphs will be reported for the primary PRO symptom scores by visits as well as change in these scores from baseline. These will also be reported for the other EORTC QLQ-C30 and EORTC QLQ-LC13 reported symptoms and scales.

Analysis of CTSQ-16

The three domains of interest (Expectations with Therapy, Feelings about Side-Effects, and Satisfaction with Therapy) will each be separately analysed using an ANCOVA stratified by race (Asian versus Non-Asian) and mutation type (Ex19del versus L858R). The results of the analyses will be presented in terms of a least squares mean together with its associated 95% profile likelihood CI. No p-values will be presented. Normality assumptions will be checked, and if a deviation from the assumptions is detected, a non-parametric analysis may be performed.

Descriptive statistics and graphs will also be reported for the CTSQ-16 three domains of interest by visit.

Tissue results based on method used for randomisation

A summary table will be produced to display the number (and percentage) of patients randomised under a central tissue test result and a local tissue test result. In addition, it will summarise the number of patients who were screened and not randomised by whether the tissue result was EGFR negative, EGFR positive where a screen failure criteria was met or EGFR unknown where a screen failure criteria was met.

In addition, a summary table will be produced to summarise the number of screened patients by the central tissue test result (EGFR positive, EGFR negative or EGFR unknown). Patients who were screened/randomised under a local test and sent a tissue sample for the central test will be included under the central test result.

A summary table with be produced to display the central test results for patients who were to be randomised under the central test (so it will be based on the screened population and include patients who were using the central tissue result as the basis for randomisation). The table will present the number of EGFR positive, EGFR negative and EGFR unknown patients. In addition, it will present the number of patients who were EGFR+ by the mutation status (Ex19del or L858R).

Sensitivity analyses for centrally confirmed EGFRm+ patients

Concordance between local testing and cobas® testing will be summarised using frequency counts, overall and by randomisation strata and positive/negative test result. Any missing results from cobas® testing will be treated as a stratum in and of itself; i.e. the mutation status results from cobas® testing will be split (for those known to have undergone testing) as EGFR positive, EGFR negative and EGFR result not available. This will allow clear comparison between local and central testing. A cross-tabulation summary of the mutation status from local EGFR testing (Positive, Negative, Not Available) versus the mutation status from central testing (Positive, Negative, Not Available) will be produced.

As sensitivity to the main analyses of PFS and ORR, analyses of these endpoints will be performed using the centrally confirmed EGFRm analysis set. In addition the main analyses of PFS and ORR will be performed for the patients with no central EGFRm+ test result available. Demographics will also be summarised (as per Section 4.3.3) using the centrally confirmed EGFRm analysis set.

Analysis of PK concentration data

All plasma concentrations of AZD9291 and its metabolites will be listed. The following summary statistics will be presented for plasma concentrations of AZD9291, its metabolites AZ5104 and AZ7550, and AZD9291:metabolite ratio (for each metabolite).

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

- The geometric mean (gmean), calculated as e^{μ}, where μ is the mean of the data on a logarithmic scale.
- Coefficient of variation (CV), calculated as $100 \times \sqrt{(exp(s^2) - 1)}$, where s is the standard deviation of the data on a log scale.
- Gmean ± standard deviation, calculated as $\pm exp(\mu + s)$.
- Arithmetic mean calculated using untransformed data.
- Standard Deviation (StdDev) calculated using untransformed data.
- Minimum.
- Median.
- Maximum.
- Number of observations (n).

Non-quantifiable (NQ) values of plasma concentrations will be handled as follows:

- If, at a given time point, 50% or less of the plasma concentrations are NQ, the gmean, CV, gmean ± StdDev, arithmetic mean and standard deviation will be calculated by substituting the limit of quantification (LOQ) for values which are NQ.
- If more than 50%, but not all, of the concentrations are NQ, the gmean, CV, gmean ± , arithmetic mean and StdDev will be reported as not calculable (NC).
- If all the concentrations are NQ, the gmean and arithmetic mean will be reported as NQ and the CV, gmean ± StdDev and StdDev as NC.
- The number of values above LLOQ will be reported for each time point along with the total number of collected values.

If data are available for less than three patients, no summary statistics other than minimum, maximum and n will be presented.

Box plots of concentration data by sample time for all available cycles for AZD9291 and metabolites (AZ5104 and AZ7550) will be displayed graphically for the PK analysis set.

4.3.8.3 Exploratory outcomes

CCI

CCI

CCI

CCI

[redacted]

CCI

[redacted]

CCI

CCI

T CCI

CCI

[redacted]

CCI

[redacted]

CCI

CCI

4.3.8.4 Analyses repeated for BICR assessment

Analyses of PFS, ORR, DoR and depth of response will be repeated using RECIST 1.1 from the BICR assessment as defined in section 4.3.8.1and 4.3.8.2. Details will be provided in the post-text table, figure and listing shells.

4.3.9 CNS efficacy analysis

The purpose of this efficacy analysis is to assess the activity of AZD9291 compared with the control arm on CNS metastases from patients in this study.

Exclusion Criteria 5 states that patient with spinal cord compression, symptomatic and unstable brain metastases, except for those patients who have completed definitive therapy, are not on steroids, have a stable neurologic status for at least 2 weeks after completion of the definitive therapy and steroids are excluded from the study.

The protocol mandated CT or MRI scans of chest and abdomen for all patients and recommends that tumour assessment coverage should include additional areas that may have tumour involvement suspected or known at screening. Therefore, brain CT / MRI were requested to be performed at screening for all patients with known or suspected brain metastasis at baseline.

Per protocol, brain metastases were recorded for RECIST tumour assessment as non-target lesions (NTLs) and assessed by the same method as per baseline assessment every 6 weeks (+/- 1 week) relative to randomisation until objective disease progression [every 12 weeks +/-

1 week following week 78 disease assessment]. All new brain/CNS lesions identified were to be recorded.

All baseline brain CT and/ or MRI scans that were submitted for BICR assessment in the study will be independently re-read by an independent neuro-radiologist in a single-read model (no double-read or adjudication will be performed). An independent neuro-radiologist will assess presence of brain disease at baseline *as measurable and / or non-measurable brain lesion*. All available subsequent brain scans (including those at unscheduled visits) will be re-read and CNS response assessed using RECIST v1.1 guidelines.

Patients with no brain scan at baseline but with a subsequent post-baseline brain scan were not included in the additional blinded central independent review.

For the CNS BICR analysis, CNS metastases identified on MRI scans and/or CT scans that measured ≥10 mm in longest diameter or ≥2 times the slice thickness/reconstruction interval, whichever was largest, were considered measurable lesions and could have been selected as target lesions (TLs). All other lesions, including all measurable lesions not selected as TLs and small lesions (i.e. <10 mm in size) or leptomeningeal disease were considered NTLs. Up to a maximum of 5 brain lesions in total were to be identified as TLs. Lesions outside the brain were not to be recorded for the purpose of this analysis. The independent neuro-radiologists received prior brain radiotherapy reports where applicable. A CNS lesion identified within sites of prior brain radiotherapy should only have been selected as a TL if no other TL was available. Further details are specified in the CNS BICR charter.

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

4.3.9.1 Definition of CNS analysis sets

CNS Full Analysis Set

The CNS Full analysis set (cFAS) will be a subset of the FAS population. It will include all patients who undertook a brain scan in the screening/baseline period, had their scan sent for independent neuro-radiology review and were identified by that review as having non-measurable and/or measurable brain disease at baseline (i.e. at least one non-measurable and/or one measurable brain lesion noted at baseline).

CNS Evaluable for Response

The CNS evaluable for response set (cEFR) will be a subset of the cFAS. It will include all patients who undertook a brain scan in the screening/baseline period, had their scan sent for independent neuro-radiology review and were identified by that review as having measurable brain disease at baseline (i.e. at least one measurable brain lesion noted at baseline).

In addition, a subgroup will be defined based on the study FAS analysis set. Patients will be classified as having a history and/or presence of a CNS lesion(s) at baseline (CNS MTS). To classify patients to a CNS MTS status at baseline, the eCRF will be used.

The programmatic definition to identify a patient with CNS MTS status of yes is shown below.

A patient is classified as CNS MTS status = yes if

a) FA.FAORRES=LOCADMET (where it is = Metastatic) and FA.FALOC=Brain/CNS

and/or

b) CM.CMCAT='CANCER THERAPY' and CM.CMTRT='Radiotherapy'. We will look for any text of 'BRAIN' or other appropriate brain-specific radiotherapy terms, based on AZ study medic review in the variable CM.CMLOC and where CMSTRF = "BEFORE"

and/or

c) A patient has a record with MH.MHDECOD matching, but not limited to (identified by AZ study medic review): Brain tumour operation, Nervous system neoplasm surgery, Neurosurgery, Metastases to central nervous system

   or if a subject has a record with CM.CMLOC matching, but not limited to the following terms (identified by AZ study medic review): left parietal brain, left temporal lobe, left temporal lobe, pan encephalic, right frontal vertex lesion, temporal, temporodorsal right, whole brain, whole-brain radiation.

Otherwise a patient will be considered with no CNS MTS at study entry (CNS MTS status=no).

The terms listed above may be extended when reviewed the AZ study medic and will be dependent on the data observed. This flag will be assigned prior to unblinding. Any scan data considered will be based on the investigator assessment.

4.3.9.2 Analysis Methods

Within this section of the SAP, if an analysis is using data from the blinded independent neuro-radiologist review, it is noted as CNS BICR. If an analysis is referring to data from the study blinded central independent review, it is noted as study BICR.

CNS BICR

A summary table will be produced to present the number of subjects included in the CNS BICR cFAS and cEFR analysis sets by treatment group.

Demographic and baseline subject characteristics will be summarised for the CNS BICR cFAS and cEFR analysis sets by treatment group. The following will be presented by total counts and percentages of subjects for the categorical variables of:

- Age
- Age (grouped as <50, ≥50-<65, ≥65)
- Sex
- Race
- Ethnic group
- EGFR mutation type (e.g. Exon 20, T790M, Exon 19 deletion, L858R)
- Overall disease classification (metastatic, locally advanced, both)
- Site of disease
- Baseline TL size (mean and categories: <40, 40-79, 80-119, ≥120mm)
- Histology
- WHO performance status (0/1)
- Prior radiotherapy to the brain (Y/N) and if yes, by ≤ 6 months or > 6 months prior to randomization
- Number of CNS lesions at baseline (e.g. 1-3, >3)

CNS BICR PFS

The cFAS analysis set will be used for this analysis. A summary of the number and percentage of patients experiencing a CNS PFS event, and the type of event (CNS progression or death) will be provided along with median, quartiles and 95% CIs.

CNS PFS will be analysed using a KM analysis. Median PFS (calculated from the KM plot, with 95% CIs), and the percentage PFS at 6, 12 and 18 months will be summarised.

The analysis method will be per the primary PFS analysis but without stratification.

A KM plot of CNS PFS will be presented.

In addition, the CNS PFS analysis above will be repeated based on the study FAS and will also include patients with no evidence of CNS disease at baseline. This will include patients who had a brain scan at baseline which indicated no lesions were present (non-target or target) and also patients who did not have a brain scan. For these patients, data from the study BICR will be used (as new CNS lesions can be reported).

For patients who did not have a brain scan at baseline, the assumption is that the patient is free of disease in the brain. For these patients the event of interest would be the occurrence of a new lesion in the brain or death due to any cause. Censoring would occur as defined in Section 3.2.3.

A subgroup analysis will be conducted comparing PFS between treatments by number of CNS lesions at baseline (1-3 and >3). For this subgroup, a Cox proportional hazards model will be fitted with the following factors:

- Treatment (AZD9291 vs. control arm)
- Number of CNS lesions at baseline (1-3 vs. >3)
- Treatment × number of CNS lesions interaction term

CNS BICR ORR

A summary of the best objective CNS response will be presented by treatment group. This is the best objective response in the CNS. It does not consider response data for overall RECIST assessment.

CNS ORR will be analysed using a logistic regression.

The CNS ORR will be compared between treatments using logistic regression models. The results of the analysis will be presented in terms of an odds ratio together with its associated 95% profile likelihood CI.

The CNS ORR analysis will be repeated for confirmed CNS BICR ORR. A response (CR and PR) is considered confirmed if it is maintained on the following scan, performed at least 4 weeks after the criteria for response were first met.

The CNS ORR analysis (unconfirmed) will be repeated by the subgroup of prior radiotherapy to the brain (radiotherapy to the brain within six months of randomisation vs. more than 6 months prior to randomisation vs. no radiotherapy to the brain).

The analysis will be conducted on the cFAS and cEFR analysis sets.

Competing risk analysis

The cFAS analysis set will be used for this analysis. The analysis will need to combine data from CNS BICR and study BICR. Study BICR will contribute progressions outside of the CNS whilst CNS BICR will contribute progressions in the CNS.

Once a patient progress, no further scans are required per protocol. Therefore, patients who progress outside of the CNS 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 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 patients who had documented radiological progression in the CNS at the same overall visit assessment as progression outside of the CNS

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)

Patients 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 patient will be assigned a value of 0 to 3, with 0 presenting patients 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 patient who is censored in the analysis, the censoring time will be the latest of the dates contributing to a particular overall visit assessment. For patients experiencing an event, the event time will be the first occurrence of an 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 6, 12 and 18 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 conditional probability of CNS progression at time t (6, 12 and 18 months) will also be provided for each treatment group. This is the probability of observing a CNS progression, conditional on the patient not experiencing a competing risk by time t.

CNS BICR - Reason for CNS progression

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

For patients with CNS progression due to a new CNS lesion, a summary will be provided to show the extent of CNS disease at baseline (i.e. it will summarise whether the patient had target and/or non-target lesions at baseline) and the response of target/non-targets lesions at the time of the appearance of the new lesion.

The cFAS analysis set will be used for this analysis.

CNS BICR - Duration and Onset of CNS Response

The cEFR analysis set will be used for this analysis. DoR (months) in responding patients (those patients who achieve a CR or PR) will be summarised using the median and 95% CI. The median will be calculated using the KM method. The number and percentage of responding patients remaining in response at >3; >6; >9; >12 months will be summarised. Additionally, descriptive data will be provided for the duration of response in responding subjects, including associated KM curves.

In order to analyze the duration of response (DoR) between arms the Expected Duration of Response (EDoR) will be derived for each treatment group (Ellis, et al., 2008). The EDoR is the product of the proportion of subjects responding to treatment and the mean DoR in responding subjects, and provides an estimate based on the cEFR analysis set.

Treatments will be compared by calculating the ratio of EDoRs using the same probability distribution for duration of response in responding subjects selected for the secondary endpoint analysis.

The analysis of DoR will weight each stratum inversely proportional to the within stratum variance of the log of the ratio of EDoRs. The mean DoR will also be estimated, and the null hypothesis that the distribution of the ratio of EDoRs is 1 will be tested and a p-value and 95% CI provided.

The time to onset of response from randomisation will also be summarised using the median calculated using the KM method. The number and percentage of responding patients with an onset by 6, 12, 18 and 24 weeks will be summarised.

The above analysis will be repeated for confirmed responses.

CNS BICR Disease control rate (DCR)

The cEFR analysis set will be used for this analysis. DCR will be analysed using a logistic regression. 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.

Best percentage change in baseline in CNS target lesions

The cEFR analysis set will be used for this analysis. The effect of AZD9291 on best percentage change in CNS tumour size will be estimated from an analysis of variance (ANOVA) model. The number of subjects, unadjusted mean and lsmeans for each treatment group will be presented, together with the difference in lsmeans, 95% CI and corresponding p-value.

In addition, a waterfall plot per treatment group will be produced.

Biological agreement between CNS ORR and overall RECIST ORR

An analysis to evaluate the concordance between CNS BICR response and overall study BICR response by RECIST in patients with CNS metastases will be performed (cFAS and cEFR analysis set), including a summary table to display the number of CNS responders/non-responders by their overall RECIST response.

An analysis to evaluate the concordance between CNS BICR response and non-CNS study BICR response by RECIST in patients with CNS metastases will be performed (cFAS analysis set), including a summary table to display the number of CNS responders/non-responders by their non-CNS RECIST response.

Concomitant radiotherapy to the brain

A summary table will be produce to show the number of patients who in addition to study treatment received radiotherapy to the brain prior to documented radiological progression.

4.3.9.3 CNS metastases status at study entry analysis

All analysis will be conducted using the study FAS population, grouped by presence or absence of CNS MTS as reported by investigator in eCRF at study entry.

A summary table will be produced to present the number of subjects by group (CNS MTS status Y/N by treatment group).

Demographic and baseline subject characteristics will be summarised by group (CNS MTS Y/N by treatment group). The following will be presented by total counts and percentages of subjects for the categorical variables of:

- Age
- Age (grouped as <50, ≥50-<65, ≥65)
- Sex
- Race
- Ethnic group

- EGFR mutation type (e.g. Exon 20, T790M, Exon 19 deletion, L858R)
- Overall disease classification (metastatic, locally advanced, both)
- Site of disease
- Baseline TL size (mean and categories: <40, 40-79, 80-119, ≥120mm)
- Histology
- WHO performance status (0/1)
- Prior radiotherapy to the brain (Y/N) and if yes, by ≤ 6 months or > 6 months prior to randomisation

ORR by investigator

A summary of the best objective response will be presented (CNS MTS status Y/N) by treatment group.

ORR will be analysed using a logistic regression.

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.

A subgroup analysis will be conducted comparing ORR between treatments by CNS MTS Y/N status. For this subgroup, a logistic regression model will be fitted with the following factors:

- Treatment (AZD9291 vs. control arm)
- CNS MTS status at baseline (Y vs. N)
- Treatment * CNS MTS status interaction term

PFS by investigator assessment

A summary of the number and percentage of patients experiencing a PFS event, and the type of event (progression or death in the absence of progression) will be provided along with median, quartiles and 95% CIs for each group (CNS MTS status by treatment group).

A subgroup analysis will be conducted comparing PFS between treatments by CNS MTS status. For this subgroup, a Cox proportional hazards model will be fitted with the following factors:

- Treatment (AZD9291 vs. control arm)
- CNS MTS status at baseline (Y vs. N)

- Treatment × CNS MTS status interaction term

A KM plot of PFS will be presented by group (CNS MTS status by treatment group) on one plot.

Reason for progression by investigator assessment

A summary table will be included to present the number/percentage of patients with RECIST progression by whether there was documented progression of a CNS lesion. The summary table will indicate whether the progression was of a non-target or new lesion (any brain lesions identified at baseline were only assessed as non-target as per RECIST v1.1 protocol guidelines). The table will present the data by group (CNS MTS status by treatment group).

If the progression was due to a CNS lesion, a summary will be presented to indicate if the progression was solely due to a CNS lesion or if other target/non-target/new lesions also indicated progression.

The number of patients reporting a new lesion in the CNS, particularly for those with no evidence of CNS involvement at study entry will not provide a complete picture. These patients may not necessarily received a brain scan post-baseline/at the time of systemic progression and therefore this could be an underestimate of patients reporting the presence of a new CNS lesion. Those patients who did have a brain scan would have been those patients with suspected CNS lesions. Asymptomatic patients without clinical suspicion of CNS lesions were not required to undertake regular scans per protocol however without the presence of a scan, the presence of CNS lesion cannot be ruled out.

PFS by study BICR

A summary of the number and percentage of patients experiencing a PFS event, and the type of event (progression or death in the absence of progression) will be provided along with median, quartiles and 95% CIs for each group (CNS MTS status by treatment group).

A subgroup analysis will be conducted comparing PFS between treatments by CNS MTS status. For this subgroup, a Cox proportional hazards model will be fitted with the following factors:

- Treatment (AZD9291 vs. control arm)
- CNS MTS status at baseline (Y vs. N)
- Treatment × CNS MTS status interaction term

A KM plot of PFS will be presented by group (CNS MTS status by treatment group).

Reason for progression by study BICR

A summary table will be included to present the number/percentage of patients with RECIST progression by whether there was documented progression of a CNS lesion. The summary

table will indicate whether the progression was of a non-target or new lesion (any brain lesions identified at baseline were only assessed as non-target as per RECIST v1.1 study protocol guidelines). The table will present the data by group (CNS MTS status by treatment group).

If the progression was due to a CNS lesion, a summary will be presented to indicate if the progression was solely due to a CNS lesion or if other target/non-target/new lesions also indicated progression.

The number of patients reporting a new lesion in the CNS, particularly for those with no evidence of CNS involvement at study entry will not provide a complete picture. These patients may not necessarily received a brain scan post-baseline/at the time of systemic progression and therefore this could be an underestimate of patients reporting the presence of a new CNS lesion. Those patients who did have a brain scan would have been those patients with suspected CNS lesions. Asymptomatic patients without clinical suspicion of CNS lesions were not required to undertake regular scans per protocol however without the presence of a scan, the presence of CNS lesion cannot be ruled out.

Duration and Onset of Response by investigator assessment

Duration of response in responding patients will be summarised using the median and 95% CI. The median will be calculated using the KM method. The number and percentage of responding patients remaining in response at >3; >6; >9; >12 months will be summarised (CNS MTS status by treatment group). Additionally, descriptive data will be provided for the duration of response in responding subjects, including associated KM curves.

In order to analyse the duration of response (DoR) between arms the Expected Duration of Response (EDoR) will be derived for each treatment group (Ellis, et al., 2008). The EDoR is the product of the proportion of subjects responding to treatment and the mean DoR in responding subjects, and provides an estimate based on the ICD evaluable for response population.

The analysis of DoR will weight each stratum inversely proportional to the within stratum variance of the log of the ratio of EDoRs. The mean DoR will also be estimated, and the null hypothesis that the distribution of the ratio of EDoRs is 1 will be tested and a p-value and 95% CI provided.

The time to onset of response from randomisation will also be summarised using the median calculated using the KM method. The number and percentage of responding patients with an onset by 6, 12, 18 and 24 weeks will be summarised.

4.3.10 Safety and tolerability

The safety analysis set will be used for all safety and tolerability tables, figures and listings except were expressly noted.

4.3.10.1 Adverse events and serious adverse events

All AEs, both in terms of MedDRA PT and CTCAE grade, will be listed and summarised descriptively by count (n) and percentage (%) for each treatment arm. MedDRA will be used for coding. Missing coding terms should be listed and summarised as "Not coded".

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

The summary tables will include all AEs that occurred after the start of treatment up until the end of the 28 day follow-up period, or the day before the first administration of crossover treatment, whichever is sooner. The 28 day follow-up period will be defined as 28 days following discontinuation of treatment.

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

Summary information (the number and percent of patients by treatment) by SOC and PT will be tabulated for:

- All AEs
- All AEs causally related to study medication
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related to treatment
- AEs with outcome of death
- AEs with outcome of death causally related to treatment
- AEs leading to dose reduction and interruption (separately)
- All SAEs
- All SAEs causally related to study medication
- AEs leading to discontinuation of treatment
- AEs leading to discontinuation of treatment, causally related to treatment
- OAEs

- OAEs causally related to treatment

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

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

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

A separate summary of AEs occurring more than 28 days after discontinuation of AZD9291 (where reported) as well as those occurring prior to treatment will be produced. These events will not be included in AE summaries.

4.3.10.2 Adverse events of special interest

Summary tables of AEs of special interest will be produced. The number (%) of patients experiencing any of the specified terms will be presented overall and by maximum CTCAE grade. Additional summaries of time to onset of first AE for each grouped term and each preferred term within it; time to onset of first CTCAE grade three or higher and duration of AEs of special interest will be produced. In addition, further summary tables from the AEs section listed above will be repeated for grouped AEs of special interest.

Life table and prevalence plots for rash and diarrhoea will also be produced.

The management of AEs of diarrhoea and rash will be summarised in terms of medication taken, dose modification and proportion of time on treatment.

Data collected on the specific eCRF for skin reactions will also be listed and summarised to further assess tolerability of specific events. Summaries will be by preferred term group for patient level and event level, overall, by effect and by CTCAE grade, for treatment-emergent skin reactions only. The number of skin reactions per patient will also be summarised.

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

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

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

4.3.10.3 Deaths

A summary of deaths will be provided with number and percentage of patients, categorised as:

- Related to disease under investigation,
- AE outcome=death,
- Both related to disease under investigation and with AE outcome=death,
- AE with outcome=death ≥28 days after last treatment dose,
- Deaths ≥28 days after last treatment dose, unrelated to AE or disease under investigation, and
- Patients with unknown reason for death.

A corresponding listing will also be produced.

4.3.10.4 Laboratory evaluations

All laboratory data recorded in the eCRF will be listed. If any additional analytes to those in Table 3 are also recorded then these will be listed only.

All values will be classified as low (below range), normal (within range) and high (above range) based on project-specific reference ranges. As applicable, values will be converted to standard units and will be graded using CTCAE v4.0.

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

Plots of both the maximum post-baseline alanine aminotransferase (ALT) and aspartate aminotransferase (AST) versus the maximum post-baseline total bilirubin, expressed as multiples of their upper limit of reference range will be produced. Box plots of absolute values and change from baseline for all haematology and clinical chemistry parameters will also be presented.

A pregnancy test will also be performed at screening only for women of child-bearing potential.

4.3.10.5 Vital signs (pulse and BP) and weight

Absolute values and change from baseline for pulse, BP and weight will be summarised by treatment group and visit.

4.3.10.6 Physical examination

Abnormalities identified from physical examination will be listed.

4.3.10.7 ECG

All ECG data received will be presented in data listings. ECG summaries will be presented for patients in the safety analysis set.

The following ECG parameters will be summarised (absolute values and change from baseline) by visit: QTcF, RR, PR, QRS and QT.

Box plots for observed ECG parameters and change from baseline in ECG parameters over time will be presented. Shift plots of the value corresponding to the maximum absolute change from baseline versus the baseline value for QTcF, with reference lines for 450 ms, ±30 ms and ±60 ms change, will be presented.

QTc outliers are defined as QTcF values following dosing that are greater than 450 ms or are increases from baseline greater than 30 ms. QTcF outliers will be highlighted in the data listings and summarised using the following categories:

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

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

4.3.10.8 Left ventricular ejection fraction

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

LVEF outliers are defined as LVEF values following dosing that are

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

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

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

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

4.3.10.9 WHO performance status

WHO performance status will be listed and summarised as frequency counts by treatment and visit.

4.3.10.10 Ophthalmologic assessment

Ophthalmologic assessment will be listed by treatment and visit.

4.3.10.11 Analyses for crossover subjects

Summary tables from the AEs sections described above, including causally related AEs, AEs leading to death, AEs leading to dose modification, serious AEs and AESIs, will be repeated for crossover subjects only.

The AE summary tables for crossover subjects will include all AEs that occurred after the start of crossover treatment up until the end of the 28 day follow-up period. The 28 day follow-up period will be defined as 28 days following discontinuation of treatment.

In addition to AE summaries, summaries of clinical laboratory parameters, vital signs and ECG data will be repeated for crossover subjects only.

5 INTERIM ANALYSES

No interim analysis will be performed.

6 CHANGES OF ANALYSIS FROM PROTOCOL

Clinically meaningful changes in EORTC QLQ-C30 and EORTC QLQ-LC13 have been removed. The analyses of symptom improvement rate and time to symptom deterioration have therefore been removed. The primary PRO symptoms will be analysed using a MMRM analysis of the change from baseline. Summary statistics and graphical displays will also be produced.

The multiple testing strategy has been updated so that the final parameter to be tested in the hierarchy is CNS PFS instead of PFS in the subgroup of T790M+ identified by a high sensitivity method.

Detailed analyses of CNS brain metastases have been added.

The alpha spending function has been revised from the fixed alpha spending approach to an O'Brien and Fleming spending function.

7 REFERENCES

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.

Berry, G., Kitchin, R. & Mock, P., 1991. A comparison of two simple hazard ratio estimators based on the logrank test. *Statistics in Medicine,* Volume 10, pp. 749-55.

Burnman, C.-F., Sonesson, C. & Guilbaud, O., 2009. A recycling framework for the contstruction of Bonferroni-based multiple tests. *Statistics in Medicine,* Volume 28, pp. 739-61.

Clopper, C. J. & Pearson, E. S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika,* Volume 26, pp. 404-413.

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

Ellis, S., Carroll, K. & Pemberton, K., 2008. Analysis of duration of response in oncology trials. *Contempory Clinical Trials,* Volume 29, pp. 456-65.

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

Glimm, E., Maurer, W. & Bretz, F., 2010. Hierarchical testing of multiple endpoints in group sequential-trials. *Statistics in Medicine,* 29(2), pp. 219-28.

Obosa, D. et al., 1998. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology,* January, 16(1), pp. 139-44.

Robins, J. M., 1993. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *ASA Proceedings of Biopharmaceutical Section,* pp. 24-33.

Robins, J. M. & Tsiatis, A. A., 1991. Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. *Commun Stat Theory Methods,* 20(8), pp. 2609-31.

Selke, T. & Siegnumd, D., 1983. Sequential analysis of the proportional hazards model. *Biometrika,* Volume 70, pp. 315-26.

Stone, A, 2010.The application of bespoke spending functions in group-sequential designs and the effect of delayed treatment switching in survival trials. Pharm Stat. 2010 Apr-Jun;9(2): 151-61.

Wu, Y.-L.et al., 2014. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol,* Volume 15, pp. 213-222.

APPENDIX 1 - ANALYSES OF DATA FROM CHINA

It is expected that data from 120 Chinese patients will be required to support a submission in China. Given the likely approval time for the protocol submission in China (assumed to be one year), the recruitment of these patients would occur towards the end of the global recruitment period, resulting in very few and insufficient PFS events from Chinese patients contributing to the primary and secondary analyses. Some patients (estimated to be approximately 30 patients) recruited in China as part of the 530 globally randomised patients will be included in the analysis of the global study. An analysis including all patients recruited in China when data are sufficiently mature (triggered from approximately 82 PFS events in the 120 Chinese patients) will be provided to the China Health authority (cFDA) only.

Definition of China analysis sets

China patients are patients enrolled in sites located in China mainland and declaring themselves as Chinese. Asian patients are patients enrolled in sites in Asia countries and declaring themselves as Asian.

China-only full analysis set

The China-only Full analysis set will include all patients randomised in China. This includes all patients randomised in China as part of the global study and all additional patients recruited in mainland China after global recruitment is completed.

The China-only Full analysis set will be used for all China-only efficacy analyses and treatment groups will be compared on the basis of randomised study treatment, regardless of the treatment actually received.

China-only safety analysis set

The China-only safety analysis set will consist of all patients randomised in China who received at least one dose of study treatment and for whom post-dose data are available.

China-only pharmacokinetic analysis set

The China-only Pharmacokinetic Analysis Set is defined as patients in the China-only FAS who have at least one evaluable PK concentration.

Primary and secondary variables for China analysis

All efficacy, safety, PRO and PK variables will be derived in the same way as detailed in Section 3.

Analysis methods for China

All analyses detailed in Section 4 will be repeated for the patients randomised in China using the analysis sets described in Appendix 1. All statistical analyses will be considered exploratory and only performed if sufficient numbers of events or patients are available, otherwise descriptive statistics only will be presented. No adjustment for multiplicity will be made and so the procedure for hierarchical testing and splitting the alpha detailed in Section 4.2.2 will not be followed.

Statistical analyses will only include a stratification variable for mutation type (Ex19del versus L858R), and will not include race (Asian versus Non-Asian).

The analyses will be performed when the PFS data from the China patients is of similar maturity to when the global analysis of PFS will be conducted; i.e. approximately 82 PFS events out of the 120 China patients. A further analysis may be performed when the OS data is more mature.

APPENDIX 2 - TUMOUR RESPONSE EVALUATION

Schedule of evaluation

Baseline assessments should encompass all areas of known predilection for metastases in the disease under evaluation and should additionally investigate areas that may be involved based on signs and symptoms of individual patients and should be performed no more than 28 days before the start of study treatment (refer to Study Plan and section 5.1 from the study protocol). Follow-up assessments will be performed every six weeks (± 1 week) for the first 18 months, and then every 12 weeks (± 1 week) after randomisation until objective disease progression as defined by RECIST 1.1 even if a patient discontinues treatment prior to progression or receives other anti-cancer treatment. Any other sites at which new disease is suspected should also be adequately imaged at follow-up.

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

Target lesions (TL)

Documentation of target lesions

A maximum of five measurable lesions, with a maximum of two lesions per organ (including lymph nodes), representative of all lesions involved should be identified as TL at baseline. Target lesions should be selected on the basis of their size (longest diameter for non-nodal lesions or short axis for nodal lesions), but in addition should be those that lend themselves to reproducible repeated measurements. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

The site and location of each TL should be documented as well as the longest diameter for non-nodal lesions (or short axis for lymph nodes). All measurements should be recorded in millimetres. At baseline the sum of the diameters for all TL will be calculated and reported as the baseline sum of diameters. At follow-up visits the sum of diameters for all TL will be calculated and reported as the follow-up sum of diameters.

Special cases:

- For TL measurable in two or three dimensions, always report the longest diameter. For pathological lymph nodes measurable in two or three dimensions, always report the short axis.

- If the CT/MRI slice thickness used is > 5mm, the minimum size of measurable disease at baseline should be twice the slice thickness of the baseline scan.

- If a lesion has completely disappeared, the longest diameter should be recorded as 0 mm.
- If a TL splits into two or more parts, then record the sum of the diameters of those parts.
- If two or more TL merge then the sum of the diameters of the combined lesion should be recorded for one of the lesions and 0 mm recorded for the other lesion(s).
- If a TL is believed to be present and is faintly seen but too small to measure, a default value of 5mm should be assigned. If an accurate measure can be given, this should be recorded, even if it is below 5mm.
- If a TL cannot be measured accurately due to it being too large, provide an estimate of the size of the lesion.
- When a TL has had any intervention; e.g. radiotherapy, embolisation, surgery etc., during the study, the size of the TL should still be provided where possible.

Evaluation of target lesions

This section provides the definitions of the criteria used to determine objective tumour visit response for TL.

Table 5 Evaluation of target lesions

CR	Disappearance of all target lesions since baseline. Any pathological lymph nodes selected as target lesions must have a reduction in short axis to < 10 mm.
PR	At least a 30% decrease in the sum of the diameters of TL, taking as reference the baseline sum of diameters
SD	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
PD	At least a 20% increase in the sum of diameters of target lesions, 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.
NE	Only relevant if any of the target lesions were not assessed or not evaluable 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 target lesion response

Evaluation of non-target 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.

Table 6 Evaluation of non-target lesions

CR	Disappearance of all non-target lesions since baseline. All lymph nodes must be non-pathological in size (< 10 mm short axis).
Non CR/Non PD	Persistence of one or more NTL
PD	Unequivocal progression of existing non-target lesions. 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.
NE	Only relevant when one or some of the non-target lesions were not assessed and, in the Investigator's opinion, they are not able to provide an evaluable overall non-target lesion assessment at this visit. Note: For patients without target lesions at baseline, this is relevant if any of the non-target lesions were not assessed at this visit and the progression criteria have not been met.

To achieve 'unequivocal progression' on the basis of non-target lesions, there must be an overall level of substantial worsening in non-target disease such that, even in presence of SD or PR in target lesions, the overall tumour burden has increased sufficiently to merit discontinuation of therapy. A modest 'increase' in the size of one or more non-target lesions is usually not sufficient to qualify for unequivocal progression status.

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.

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

If a new lesion is equivocal, for example because of its small size, the treatment and tumour assessments should be continued until the new lesion has been confirmed. If repeat scans

confirm there is a new lesion, then the progression date should be declared using the date of the initial scan.

Symptomatic deterioration

Symptomatic deterioration is not a descriptor of an objective response: it is a reason for stopping study therapy.

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

Evaluation of overall visit response

The overall visit response will be derived using the algorithm shown in Table 7.

Table 7 Overall visit response

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

Evaluation of best overall response

The best overall response for each patient will be determined from the sequence of overall visit responses according to the following rules:

- CR = at least one visit response of CR before progression.
- PR = at least one visit response of PR before progression (and not qualifying for a CR).

- SD = at least one SD assessment >6 weeks after randomisation (allowing for 7 day visit window, so at least 35 days after randomisation) and before progression (and not qualifying for CR or PR).

- PD = progression (and not qualifying for CR, PR or SD).

- NE = all other cases (i.e. not qualifying for CR, PR or PD and without SD after more than 35 days).

However, any visit response which occurred after a further anti-cancer therapy was received will not be included in the calculation of best overall response.