AstraZeneca

Statistical Analysis Plan

Study Code	D5160C00003
Edition Number	4
Date	07 July 2016

A Phase III, Open Label, Randomized Study of AZD9291 versus Platinum-Based Doublet Chemotherapy for Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer whose Disease has Progressed with Previous Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor Therapy and whose Tumours harbour a T790M mutation within the Epidermal Growth Factor Receptor Gene (AURA3)

A Phase III, Open Label, Randomized Study of AZD9291 versus Platinum-Based Doublet Chemotherapy for Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer whose Disease has Progressed with Previous Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor Therapy and whose Tumours harbour a T790M mutation within the Epidermal Growth Factor Receptor Gene (AURA3)

Study Statistician PPD _____ 07/07/2016
Date

A Phase III, Open Label, Randomized Study of AZD9291 versus Platinum-Based Doublet Chemotherapy for Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer whose Disease has Progressed with Previous Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor Therapy and whose Tumours harbour a T790M mutation within the Epidermal Growth Factor Receptor Gene (AURA3)

Global Product Statistician PPD 07/07/2016

Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ANCOVA	Analysis of covariance
ATC	Anatomical Therapeutic Chemical
AUC 5	Area under the plasma concentration-time curve 5 mg/ml per minute
$AUC_{(0-24)}$	Area under the curve (plasma concentration / time curve from zero to 24 hours)
AUC_{ss}	Area under the plasma concentration-time curve at steady state
BICR	Blinded Independent Central Review
BP	Blood pressure
CSF	Cerebrospinal fluid
CI	Confidence interval
Cl_{ss}/F	Plasma clearance at steady state
C_{max}	Maximum plasma concentration
CMET	Proto-oncogene encoding Hepatocyte Growth Factor Receptor
CR	Complete response
CSP	Clinical study protocol
CSR	Clinical study report
C_{ss}	Concentration at study state
CT	Computerized tomography scan
CTCAE	Common Terminology Criteria for Adverse Event
ctDNA	Circulating tumor DNA
DCR	Disease control rate
dECG	Digital electrocardiogram
DNA	Deoxyribonucleic acid
DoR	Duration of response
ECG	Electrocardiogram
eCRF	Electronic case report form
EDoR	Expected duration of response
EGFR	Epidermal growth factor receptor
EGFRm+	EGFR mutation positive

Abbreviation or special term	Explanation
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Core 30 items
EORTC QLQ-LC13	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Lung Cancer 13 items
EQ-5D-5L	European Quality of Life - 5 Dimensions - 5 Levels
EuroQoL	European Quality of Life
FAS	Full analysis set
FDA	Food and Drug Administration
HER2	Human epidermal growth factor receptor 2
HP	Haybittle-Peto
HR	Hazard ratio
HRR	Hazard ratio ratio
HRQoL	Health Related Quality of Life
IP	Investigational product
IPCW	Time Inverse Probability of Censoring Weighting
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
KM	Kaplan-Meier
L858R	An amino acid substitution at position 858 in EGFR, from a Leucine (L) to an Arginine (R)
LD	Longest diameter
LDH	Lactate dehydrogenase
LE	Local evaluation
LLoQ	Lower limit of quantification
LVEF	Left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
MUGA	Multi-gated acquisition scan
NE	Not evaluable
NSCLC	Non-small cell lung cancer
NTL	Non-target lesion
OAE	Other significant adverse event
ORR	Objective response rate

Abbreviation or special term	Explanation
OS	Overall survival
p.o.	Per os, by mouth, orally
PD	Progression of disease
PFS	Progression Free Survival
PID	Percentage intended dose
PFS2	Time from randomization to second progression
PK	Pharmacokinetics
PR	Partial response
PRO	Patient reported outcome
PT	Preferred term
QD	Once daily
RAC	Ratio of AUCss after multiple dosing to AUC(0-24) after single dose
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumors
RPSFT	Rank Preserving Structural Failure
SAE	Serious adverse event
SD	Stable disease
SLD	Sum of longest diameter
SOC	System organ class
T790M	An amino acid substitution at position 790 in EGFR, from a Threonine (T) to a Methionine (M)
T790M-	T790M mutation negative
T790M+	T790M mutation positive
TFST	Time to first subsequent therapy or death
TKI	Tyrosine kinase inhibitor
TL	Target lesion
t_{max}	Time to C_{max}
$t_{ss\ max}$	Time to $C_{ss\ max}$
TSST	Time to second subsequent therapy or death
ULoQ	Upper limit of quantification
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief description of change
02Jul2015	SAP updated based on protocol amendment 2
18Mar2016	SA updated based on protocol amendment 3

1. STUDY DETAILS

1.1 Study objectives

Primary objective

Objective:	Outcome Measure:
To assess the efficacy of AZD9291 compared with platinum-based doublet chemotherapy by assessment of Progression Free Survival (PFS).	PFS using investigator assessments according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.1. Sensitivity analysis of PFS using Blinded Independent Central Review (BICR).

Secondary objectives

Objective:	Outcome Measure:
To further assess the efficacy of AZD9291 compared with platinum-based doublet chemotherapy in terms of: - Objective Response Rate (ORR) - Duration of Response (DoR) - Disease Control Rate (DCR) - Overall Survival (OS) - Tumor shrinkage	ORR, DoR, DCR and tumor shrinkage using investigator assessments according to RECIST 1.1. Analysis of OS.
To assess the effect of AZD9291 compared to platinum-based doublet chemotherapy on subjects' disease-related symptoms and health related quality of life (HRQoL).	EORTC QLQ-C30: Questionnaire consisting of 30 items measuring subjects general cancer symptoms and functioning. EORTC QLQ-LC13: A complementary questionnaire measuring lung cancer symptoms
To characterize the pharmacokinetics (PK) of AZD9291, AZ5104 and AZ7550 in subjects receiving AZD9291.	Plasma concentrations of AZD9291, and metabolites AZ5104 and AZ7550. PK parameters (such as Cl_{ss}/F, $C_{ss\ min}$ and $C_{ss\ max}$ AUC_{ss}) will be derived using population PK analysis and reported separately to the Clinical Study Report (CSR). Data from this study may form part of a pooled analysis with data from other studies.

Safety objective

Objective:	Outcome Measure:
To assess the safety and tolerability profile of AZD9291 compared with platinum-based doublet chemotherapy.	- Adverse events, graded by Common Terminology Criteria for Adverse Event (CTCAE v4) - Clinical chemistry, hematology and urinalysis - Vital signs (pulse and blood pressure), physical examination, weight - Centrally reviewed digital electrocardiogram (dECG) - Multi-gated acquisition scan (MUGA) for left ventricular ejection fraction (LVEF) - World Health Organization (WHO) performance status

Exploratory objectives

Objective:	Outcome Measure:
To explore the relationship between PK and selected endpoints (which may include efficacy, safety and/or Patient Reported Outcome [PRO]), where deemed appropriate.	Correlation of PK with other primary/secondary/ exploratory endpoints in subjects treated with AZD9291. Results from such analyses will be reported separately from the CSR. Data from this study may also form part of a pooled analysis with other AZD9291 studies.
To compare the effects of AZD9291 compared with chemotherapy on post-progression outcomes	Time from randomization to second progression (PFS2) Time to subsequent therapy Time to change in symptoms (including post-progression assessments)
To further characterize the effects of AZD9291 on survival outcomes.	Assess the impact on OS of baseline potentially prognostic factors (e.g., tumor stage, performance status, sex, baseline lactate dehydrogenase [LDH]). Assess the impact on OS of subsequent treatments and other potential covariates (e.g., changes in performance status).
To compare adverse events by subject self-reporting of specific CTCAE symptoms (where applicable) between AZD9291 and chemotherapy.	Collection of approximately 28 PRO-CTCAE symptoms via an electronic device solution (in countries where language is available).

Objective:	Outcome Measure:
To compare AZD9291 treatment with chemotherapy treatment on health state utility.	The European Quality of Life - 5 Dimensions - 5 Levels (EQ-5D-5L) health state utility index will be used to derive health state utility based on subject reported data.
To compare health resource use associated with AZD9291 treatment with chemotherapy treatment.	Health resource utilization measures including hospitalization, outpatient visits, or emergency department visits.
To characterize the PK of AZD9291, AZ5104 and AZ7550 in cerebrospinal fluid (CSF).	Concentration of AZD9291, AZ5104 and AZ7550. Ratio of metabolites to AZD9291. Ratio of CSF to plasma concentration. Summaries of PK concentration data.
To collect and store deoxyribonucleic acid (DNA) for future exploratory research into genes/genetic variation that may influence PK or response to AZD9291, platinum-based doublet chemotherapy (i.e., absorption, distribution, metabolism, excretion, safety and efficacy) and/or susceptibility to/development of cancers.	Correlation of polymorphisms with variation in PK, pharmacodynamics, safety or response observed in subjects treated with AZD9291 or comparator. Data generated may be reported separately and may also form part of a pooled analysis with other AZD9291 studies.
To collect and store tumor samples and blood-based samples for potential for exploratory research into factors that may influence susceptibility to/development of non-small cell lung cancer (NSCLC)/cancer and/or response to AZD9291 (where response is defined broadly to include efficacy, tolerability or safety).	Analysis of key genetic and proteomic markers to include, but not limited to, epidermal growth factor receptor (EGFR) mutations, human epidermal growth factor receptor 2 (HER2) and proto-oncogene encoding hepatocyte growth factor receptor (CMET) expression and/or amplication. Collection of plasma samples to include, but not be limited to, extraction of circulating tumor DNA (ctDNA) for investigation of blood-borne biomarkers. The sample may be used to investigate the relationship between PK and blood-borne biomarkers. Samples may be analyzed retrospectively. Any biomarker data generated may be reported separately and may also form part of a pooled analysis with other AZD9291 studies.
To collect and store plasma for isolation of ctDNA. Extracted ctDNA will be assessed for the presence of genetic aberrations including, but not limited to, EGFR mutations (T790M, L858R, etc). These samples may be shared with a diagnostic partner for development of a ctDNA test for T790M detection.	Retrospective/real-time analysis of EGFR (and other) mutations in ctDNA from all study subjects (mandatory).

Objective:	Outcome Measure:
To collect and store residual CSF for potential exploratory research of factors that may influence development of NSCLC and/or response to AZD9291 (where response is defined broadly to include efficacy, tolerability or safety).	Collection of CSF for the investigation of PK and/or biomarkers. Samples may be analyzed retrospectively.

Results from such analyses outlined above may be reported separately from the CSR.

1.2 Study design

This is a phase III, open-label, randomized study assessing the safety and efficacy of AZD9291 (80 mg, orally [p.o.], once daily [QD]) versus platinum-based doublet chemotherapy in second-line subjects with EGFR mutation and T790M mutation positive (EGFRm+/T790M+), locally advanced or metastatic NSCLC who have progressed following treatment with an approved EGFR Tyrosine Kinase Inhibitor (EGFR-TKI) agent.

A biopsy will be needed for central testing of T790M mutation status following confirmed disease progression on first-line treatment with an EGFR-TKI.

Approximately 410 subjects, stratified by ethnicity (Asian/non-Asian) will be centrally randomized to receive either AZD9291 (80mg QD p.o.) or platinum-based doublet chemotherapy (pemetrexed 500 mg/m^2 + carboplatin area under the plasma concentration-time curve AUC 5 or pemetrexed 500 mg/m^2 + cisplatin 75 mg/m^2) on Day 1 of every 21-day cycle in a 2:1 (AZD9291: platinum-based doublet chemotherapy) ratio using the Interactive Voice Response (IVRS) or Interactive Web Response System (IWRS). The investigational site must declare (in IVRS/IWRS), prior to randomization, their choice of chemotherapy for that subject. Once 410 subjects have been randomized globally, recruitment will continue only in mainland China until approximately 50 Chinese subjects have been randomized.

Subjects should continue on investigational product (IP) AZD9291 until a treatment discontinuation criterion is met. Subjects may continue to receive AZD9291 beyond RECIST 1.1 defined progression as long as they show clinical benefit, as judged by the investigator.

Subjects can receive up to 6 cycles of pemetrexed + carboplatin/cisplatin as initial treatment.

Subjects whose disease has not progressed after four cycles of platinum-based first-line chemotherapy may receive pemetrexed maintenance therapy. Subjects who progress according to RECIST 1.1 criteria prior to completion of initial doublet chemotherapy treatment or during pemetrexed maintenance monotherapy, may continue with chemotherapy treatment as long as they are continuing to show clinical benefit, as judged by the investigator.

Subjects must be followed until evidence of RECIST 1.1 defined progression (regardless of reason for treatment discontinuation). It is important that subjects are assessed according to the intended scanning schedule to prevent the bias in analysis that can occur if one treatment group is assessed more or less often than the other.

Once subjects on the platinum-based doublet chemotherapy arm are determined to have objective radiological progression according to RECIST 1.1 by the investigator and confirmed by independent central imaging review, they will be given the opportunity to begin treatment with AZD9291 80mg, once daily. These subjects may continue treatment with AZD9291, as long as they are continuing to show clinical benefit, as judged by the investigator.

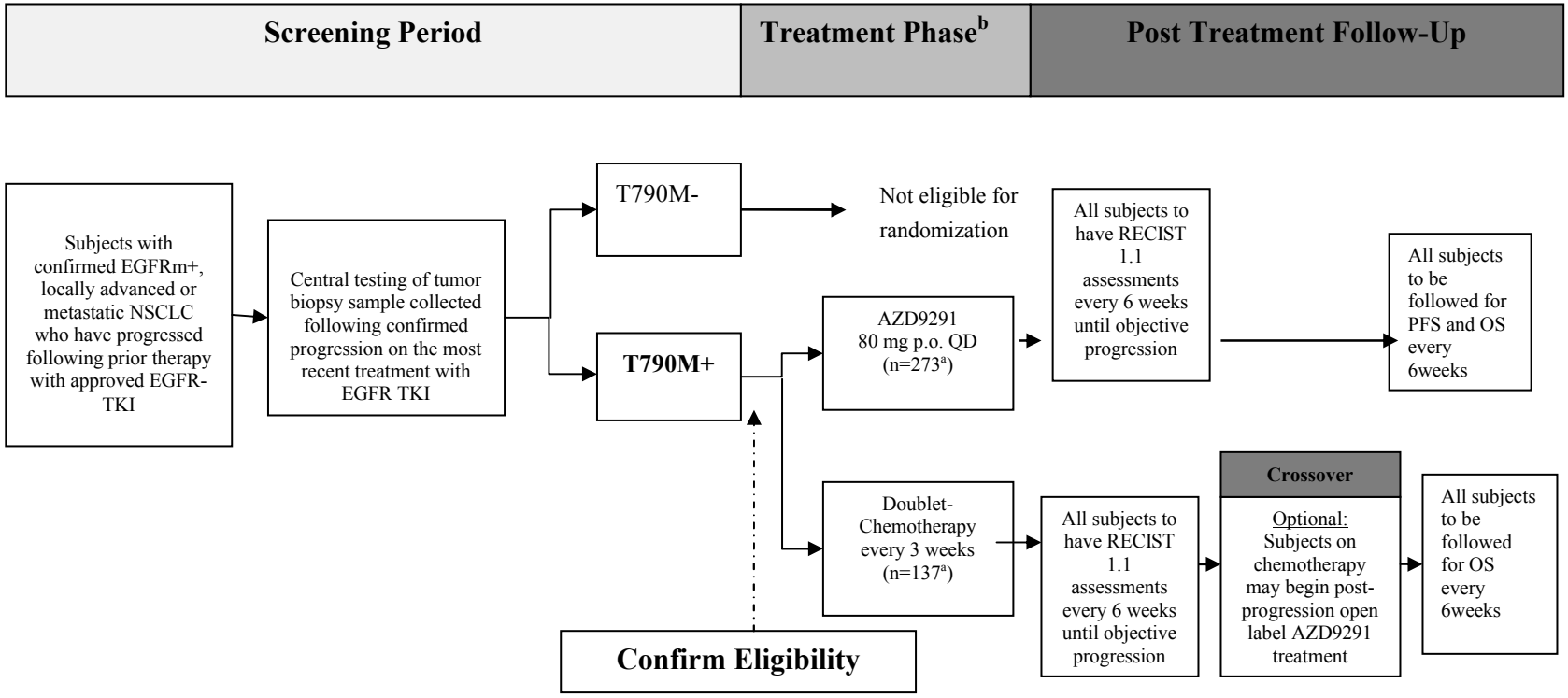
If a subject has been deemed to have objective disease progression according to Investigator tumour assessment according to RECIST 1.1, but is not confirmed by independent central imaging review, he/she is not eligible to cross-over to AZD9291 at that time. Should it be in the subjects' best interests, and only if further randomized chemotherapy is warranted, he/she may continue to receive randomized doublet chemotherapy (if initial doublet chemotherapy treatment has not yet been completed) or Pemetrexed maintenance monotherapy, sites are to submit the next scheduled tumour assessment for central imaging review according to Table 2 of the CSP.

If subjects are not eligible to cross-over to AZD9291, they will enter into the follow-up phase of the study, and other treatment options should be discussed by the investigator.

Note: Throughout this document the term "platinum-based doublet chemotherapy", refers to "platinum" as both carboplatin and cisplatin.

The overall study design is shown in Figure 1 of the clinical study protocol (CSP) and the study flow chart in Figure 1 below. The study schedule for AZD9291 subjects and for platinum-based doublet chemotherapy subjects is detailed in Tables 1 and 2 of the CSP, respectively.

Figure 1 Study flow chart



(a) Approximately 410 subjects will be randomized into the trial; from these approximately 273 subjects will be treated with AZD9291 and approximately 137 subjects will be treated with a platinum based doublet-chemotherapy.

(b) Subjects will be considered in "Treatment phase" at the time AZD9291/Doublet-Chemotherapy (study treatment) is started. Subjects will continue to receive study treatment until objective disease progression (according to RECIST 1.1). Subjects randomized to AZD9291 may continue to receive AZD9291 as long as they are continuing to show clinical benefit, as judged by the investigator, and in the absence of discontinuation criteria. Subjects randomized to the platinum-based double chemotherapy arm with objective radiological progression according to RECIST 1.1 by the investigator and confirmed by independent central imaging review will be given the opportunity to begin treatment with AZD9291, 80 mg once daily.

1.3 Number of subjects

PFS is the primary endpoint for this study. Approximately 410 subjects, stratified by ethnicity (Asian/non-Asian) will be randomized in a 2:1 ratio to the treatments as specified below:

- AZD9291 80mg QD p.o.
- Platinum-based doublet chemotherapy (pemetrexed 500mg/m^2 + carboplatin AUC5 or pemetrexed 500mg/m^2 + cisplatin 75mg/m^2)

The primary analysis of PFS will occur when at least 221 progression events have been observed out of the 410 globally randomized subjects. With 221 progression events, the study will have at least 80% power to show a statistically significant PFS at the 5% 2-sided significance level if the assumed treatment effect were hazard ratio (HR) 0.67; this translates to a 3 month improvement on an estimated median PFS of 6 months on the control arm, assuming PFS is exponentially distributed. The smallest treatment difference that would be statistically significant is a PFS HR of 0.76. Assuming 15 months non-linear recruitment, 221 PFS events are expected to occur approximately 20 months after the first subject is randomized to the study.

Three analyses of OS are planned; two interim analyses and a final analysis. The data cut-off for the first analysis of overall survival will be conducted approximately 4 months after data cut-off for the primary analysis of PFS and the second analysis will be conducted when the OS data are approximately 50% mature (approximately 205 death events). A final analysis of overall survival will be performed when the OS data are approximately 70% mature (approximately 287 death events).

In order to randomize 410 subjects (assuming a 10% screen failure rate), 455 T790M+ subjects will need to be enrolled. In order to enroll 455 T790M+ subjects, 1034 subjects will need to be screened (allowing for a 20% attrition rate and assuming 55% of subjects are T790M+).

2. ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 Full analysis set

The full analysis set (FAS) will include all randomized subjects prior to the end of global recruitment. Any subjects randomized in China, after global recruitment has ended, will not be included in the FAS.

The FAS, or subsets thereof as per 2.1.1-2, will be used for all efficacy and exploratory analyses and treatment groups will be compared on the basis of randomized treatment, regardless of the treatment actually received.

2.1.2 Safety analysis set

The safety analysis set will consist of all subjects randomized prior to the end of global recruitment who received at least one dose of randomized treatment and for whom post-dose data are available. Any subjects randomized in China only, after global recruitment has ended, will not be included in the safety analysis set.

Safety data will not be formally analyzed but summarized using the safety analysis set, according to the treatment received (i.e., erroneously treated subjects [e.g., those randomized to treatment A but actually given treatment B] will be summarized according to the treatment they actually received).

2.1.3 PK analysis set

The PK analysis set includes subjects in the FAS who have at least one measurable PK concentration and who have the relevant date, time and dosing data for this sample. The dosing data must include the dosing data for the calendar day prior to the sample day as well as the sample day.

Table 1 Summary of outcome variables and analysis sets

Outcome variable	Analysis Set
Efficacy Data: PFS, ORR, DoR, DCR, Tumor shrinkage, OS, symptoms/HRQoL	FAS
Demography	FAS
PK data	PK (subset of FAS)
Safety Data	Safety
Exploratory Data	FAS

Data on the number of subjects and duration of treatment for subjects who begin treatment with AZD9291 following centrally confirmed objective progression will also be summarized separately, as described in Section 4.2.6.

2.1.4 China extension

In order to adequately compare the efficacy and safety of AZD9291 versus chemotherapy in Chinese patients to fulfill China Food and Drug Administration (FDA) requirements, subjects recruited in China after global recruitment has ended will be combined with any subjects recruited from China prior to the end of the global recruitment. This China-only cohort will be analysed and reported separately from the CSR. These analyses will be performed when approximately 27 PFS events have been observed out of approximately 50 Chinese subjects (54% maturity, as consistent with the primary global PFS analysis).

All efficacy, safety, PRO and PK variables will be derived in the same way as detailed in Section 4. All analyses detailed in Section 4 will be repeated for the China-only cohort using the analysis sets described in this section.

All statistical analyses will be considered exploratory and only performed if sufficient numbers of events or subjects are available (e.g. > 20 PFS or OS events), otherwise descriptive statistics only will be presented. No adjustment for multiplicity will be made and so the procedure for hierarchical testing detailed in protocol Section 8.5.2.1 will not be followed. Statistical analyses will not be stratified by ethnicity (Asian versus Non-Asian).

2.1.4.1 China-only Full Analysis Set

The China-only FAS will include all subjects randomized in China. This includes all subjects randomized in China prior to the end of global recruitment and all additional subjects recruited in mainland China after global recruitment is completed.

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

2.1.4.2 China-only Safety Analysis Set

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

2.1.4.3 China-only Pharmacokinetic Analysis Set

The China-only PK analysis set is defined as subjects in the China-only FAS who have at least one evaluable PK concentration without any violation or deviation that impacts the PK analysis.

2.2 Violations and deviations

All important deviations related to the study inclusion or exclusion criteria and study conduct will be listed and summarized by randomized treatment group. None of the deviations will lead to any subjects being excluded from any of the analysis sets described in Section 2.1 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). If the deviations are serious enough to have the potential to impact the primary analysis, sensitivity analyses may be performed.

The following general categories will be considered important deviations. This list is not exhaustive and additional important deviations may be added prior to database lock.

- Informed consent procedure deviation (e.g., no informed consent signed prior to any screening procedure)
- Eligibility criteria deviation (e.g., any inclusion criteria not met or exclusion criteria met)

- Prohibited medication deviation (e.g., subject received other anticancer agents, investigational agents, or radiotherapy while on study treatment)

The categorization of these as important deviations is not automatic and will depend on duration and the perceived effect on efficacy.

In addition to the programmatic determination of the deviations above, monitoring notes or summaries will be reviewed to determine any important post-entry deviations that are not identifiable via programming, and to check that those identified via programming are correctly classified. The final classification will be made prior to database lock.

Misrandomizations in terms of errors in treatment dispensing, in addition to incorrect stratifications, will also be summarized and listed, separately to the important protocol deviations. A misrandomization is when a subject is not randomized or treated according to the randomization schedule. It includes subjects who receive no treatment whatsoever for a period of time due to errors in dispensing of medication. Note, this is not due to tolerability issues where subjects may stop taking drug.

The summary will include all subjects with a dispensing error and will also include information on how many of those subjects received at least one dose of the wrong treatment (AZD9291/platinum-based doublet chemotherapy) at any time. Subjects who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 2.1. During the study, decisions on how to handle misrandomizations will be made on an individual basis with written instruction from the study team leader and/or statistician.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Efficacy assessments

Investigator's RECIST-based assessments

From the investigator's review of the imaging scans, the RECIST tumor response data will be used to determine each subject'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. The endpoint of tumor shrinkage will be assessed from tumor size measurements based on investigator assessment.

Baseline radiological tumor assessments are to be performed no more than 28 days before the start of randomized treatment. If a baseline radiological tumor assessment is performed more than 28 days before the start of randomized treatment, then the scan will still be included as baseline but will be recorded as an important protocol deviation. Tumor assessments are then performed every 6 weeks relative to randomization until disease progression. Baseline values recorded after treatment start should not be used as the baseline assessment; although such assessments can be used in the calculation of PD. At each visit, subjects will be programmatically assigned an overall visit response of complete (CR), partial (PR), stable

disease (SD) or progressive disease (PD) depending on the status of their disease compared with baseline and previous assessments. If a subject has had a tumor assessment which cannot be evaluated, then the subject 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 CSP Appendix H 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.

Rounding of TL 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%.

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 ≥5 mm from nadir even assuming the non-recorded TLs have disappeared.

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

Lymph nodes

For lymph nodes, if the size reduces to <10 mm, these are considered non-pathological.

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

TL visit responses subsequent to CR

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., 0 mm or <10 mm for lymph nodes), the response will be set to CR irrespective of whether the criteria for PD of TL is also met (i.e., if a lymph node LD increases by 20% but remains <10 mm).

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e., 0 mm or <10 mm for 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.

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

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure a value of 5 mm 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.

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, the lesion diameter (for those lesions with intervention) will be treated as missing and scaled up as described below 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 subject will be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then if ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters will be calculated, and PR or SD then assigned as the visit response. Subjects with intervention are evaluable for

CR as long as all non-intervened lesions are 0 (or <10 mm for lymph nodes) and the lesions that have been subject to intervention also have a value of 0 recorded.

- If scaling up is not appropriate due to > 1/3 missing measurements then the visit response will be set as NE.

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

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

Scaling (applicable only for irradiated lesions/lesion intervention)

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

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

Example of scaling

Lesion	Longest diameter at nadir visit	Longest diameter at follow-up visir
1	7.2	7.1
2	6.7	6.4
3	4.3	4.0
4	8.6	8.5
5	2.5	Missing
Sum	29.3	26

Lesion 5 is missing at the follow-up visit.

The sum of lesions 1-4 at the follow-up is 26 cm. The sum of the corresponding lesions at the nadir visit is 26.8 cm.

Scale up as follows to give an estimated TL sum of 28.4cm:

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

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

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

Computerized tomography (CT) scan 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 scan and MRI this will be considered acceptable and no adjustment within the programming is needed.

BICR of RECIST based assessments

BICR of radiological imaging data will be carried out using RECIST 1.1. All radiological scans for all subjects (including those at unscheduled visits, or outside visit windows) will be provided to the BICR. The imaging scans will be reviewed by at least two independent radiologists using RECIST 1.1 criteria and will be adjudicated if required.

Up to three qualified radiologists will independently review the imaging scans in the following way. First a primary review will be performed by two independent radiologists for each patient on a timepoint by timepoint basis to give an overall tumour assessment at each timepoint using RECIST 1.1. Then a global radiology review will be performed whereby the same independent radiologists will globally assess all timepoints for a patient in the review period and adjust an overall assessment if necessary. Finally, if the overall assessment for at least one timepoint 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). These visit responses (either from the adjudicator selected reviewer, or the primary reviewer in the absence of adjudication) will be used to report BICR visit responses, BICR best overall response, BICR time to response, BICR censoring of response, BICR progression date and BICR progression censoring.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer. In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to all BICR RECIST endpoints including dates of progression, response, visit response, censoring and changes in target lesion dimensions.

All independent reviewers will be blinded to treatment. For each subject, the BICR will define the overall visit response data (CR, PR, SD, PD or NE) and the relevant scan dates for each time point (i.e., for visits where response or progression is/is not identified).

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

3.1.1 PFS

PFS is defined as the time from randomization until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy prior to progression. Subjects who have not had PD 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 subject has PD or dies after two or more missed visits, the subject will be censored at the time of the latest evaluable RECIST assessment. If the subject has no evaluable visits or does not have baseline data they will be censored at 0 days unless they die within two visits of baseline.

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 subject for PFS the subject will be censored at the latest of the dates contributing to a particular overall visit assessment

3.1.2 ORR

ORR rate is defined as the number and percentage of randomized subjects 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 complete response or partial response, which occurred after a further anti-cancer therapy was received, will not be included in the numerator for the ORR calculation.

3.1.3 DoR

DoR will be defined as the time from the date of first documented response until the date of documented PD or death in the absence of PD, the end of response should coincide with the date of PD 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 subject does not progress following a response, then the subject's duration of response will use the PFS censoring time.

3.1.4 DCR

DCR is defined as the percentage of subjects who have a best overall response of CR or PR or SD at ≥6 weeks (± 7 day visit window), prior to any PD event.

3.1.5 Tumor shrinkage

Tumor shrinkage will be assessed using RECIST 1.1 tumor response. The absolute change and percentage change from baseline in sum of tumor size at each assessment will be calculated.

Tumor size is the sum of the longest diameters (SLD) of the TLs. The percentage change in SLD at each week for which data are available will be obtained for each subject taking the difference between the SLD at each week and the SLD at baseline divided by the SLD at baseline multiplied by 100 [i.e., (week n - baseline) / baseline x 100]. The change from baseline will be obtained for each subject taking the difference between the SLD at each week and the SLD at baseline (i.e., week n - baseline).

The best percentage change in SLD from baseline or the minimum increase from baseline in the absence of a reduction from baseline based on all post-baseline assessments prior to the visit when progression is detected or start of subsequent anti-cancer therapy.

If best percentage change cannot be calculated due to missing data, 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;
- If a subject has withdrawn due to PD and has no evaluable TL data before or at PD.

3.1.6 OS

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

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.

3.1.7 HRQoL and symptoms

PROs will be assessed using the EORTC QLQ-C30 and EORTC QLQ-LC13 questionnaires. The EORTC QLQ-C30 consists of 30 questions, which can be combined to produce 5 functional scales (Physical, Role, Cognitive, Emotional, Social), 3 symptom scales (Fatigue, Pain, Nausea/vomiting), 5 individual items (dyspnea, insomnia, appetite loss, constipation, diarrhea) and a global measure of health status. The EORTC QLQ-LC13 is a lung cancer specific module comprising 13 questions to assess lung cancer symptoms (cough, hemoptysis, dyspnea and site-specific pain), treatment related side-effects (sore mouth, dysphagia, peripheral neuropathy and alopecia) and pain medication. With the exception of a multi-item scale for dyspnea, all are single items.

An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/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, respectively.

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.

For each symptom/scale, if less than 50% of the subscale items are missing, the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscale. If at least 50% of the items are missing, that subscale also will be treated as missing.

For analysis of time to symptom deterioration, the FAS will be subset to include subjects who have EORTC QLQ-LC13 baseline scores ≤90.

For analysis of symptom improvement rate, the FAS will be subset to include subjects who have a baseline score ≥10 for EORTC QLQ-C30 scales/items or EORTC QLQ-LC13 scales/items.

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

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

- fatigue (multi-item scale based on three questions: "Did you need rest; Have you felt weak; Were you tired?")
- appetite loss: one item ("Have you lacked appetite?").

Definition of clinically meaningful changes

Changes in score compared to baseline will be evaluated. A minimum clinically meaningful change is defined as a change in the score from baseline of ≥10 for scales/items (Obosa et al 1998).

For example, a clinically meaningful deterioration or worsening in chest pain (as assessed by EORTC QLQ-LC13) is defined as an increase in the score from baseline of ≥10. A clinically meaningful improvement in fatigue (as assessed by EORTC QLQ-C30) is defined as a decrease in the score from baseline of ≥10.

At each post-baseline assessment, change in symptoms/functioning from baseline will be categorized as improved, stable or worsening as shown in Table 2.

Table 2 Visit response for HRQoL and disease-related symptoms

Score	Change from baseline	Visit response
EORTC QLQ-LC13 symptom scales/items	≥+10	Worsened
	≤-10	Improved
	Otherwise	Stable
EORTC QLQ-C30 symptoms scales/items	≥+10	Worsened
	≤-10	Improved
	Otherwise	Stable
EORTC QLQ-C30 functional scales and Global health status	≥+10	Improved
	≤-10	Worsened
	Otherwise	Stable

Compliance

Summary measures of compliance at each scheduled assessment will be derived for the EORTC QLQ-C13 and QLQ-C30 score. 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, 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.

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

3.1.7.1 Time to symptom deterioration

For each of the symptoms scales/items in EORTC QLQ-LC13, time to symptom deterioration will be defined as the time from randomization until the date of first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy prior to symptom deterioration. Death will be included as an event only if the death occurs within two visits of the last PRO assessment where the symptom change could be evaluated.

Subjects whose symptoms (as measured by EORTC QLQ-LC13) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms progress after two or more missed PRO assessment visits or the subject dies after two or more missed PRO assessment visits, the subject will be censored at the time of the last PRO assessment where the symptom could be evaluated. If a subject has no evaluable visits or does not have baseline data they will be censored at Day 1. The population for analysis of time to symptom deterioration will include a subset of the FAS who have EORTC QLQ-LC13 or EORTC QLQ-C30 baseline scores ≤90.

3.1.7.2 Symptom improvement rate

The symptom improvement rate will be defined as the number (%) of subjects with two consecutive assessments at least 18 days apart (i.e., 21 days allowing a visit window of 3 days), which showed a clinically meaningful improvement (a decrease from baseline score ≥10 for EORTC QLQ-LC13 scales/items and EORTC QLQ-C30 scales/items) in that symptom from baseline. The denominator will consist of a subset of the FAS who have a baseline symptom score ≥10.

3.1.7.3 PFS2

PFS2 is defined as the time from the date of randomization to the earliest of the progression event subsequent to that used for the primary variable PFS or death. Subjects who are alive and for whom a second PD has not been observed will be censored at their last assessment where the subject has not had a second progression or death. If a subject dies without a progression event, the subject's PFS and PFS2 events would be equivalent.

For subjects with no PFS and no PFS2, the PFS2 censor date will be same as PFS censor date, with the following exceptions:

- If there is no PFS event but a PFS2 event has been recorded, the PFS2 event should be used in the PFS2 analysis even though the subject will be censored in PFS analysis.

- If there is no PFS event but the PFS2 form has been completed as no event with a date, then use this date as the censor date (rather than PFS censor date).

3.1.7.4 TFST

TFST is defined as the time from the date of randomization to the earlier of the date of anti-cancer therapy start date following study treatment discontinuation, or death. Any subject 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). If a subject terminates the study for a reason other than death before first subsequent anti-cancer therapy, the subject will be censored at the termination date.

3.1.7.5 TSST

TSST is defined as the time from the date of randomization to the earlier of the date of second subsequent anti-cancer therapy start date following study treatment discontinuation, or death. Any subject not known to have died at the time of the analysis and not known to have had a second further intervention of this type will be censored at the last known time to have not received second subsequent therapy (i.e., the last follow-up visit where this was confirmed). If a subject terminates the study for a reason other than death before the second subsequent anti-cancer therapy, the subject will be censored at the termination date.

3.1.8 PRO-CTCAE symptoms

The PRO-CTCAE questionnaire will be used to derive subject reporting of CTCAE 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.

3.1.9 Health state utility

The EQ-5D-5L index comprises 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). For each dimension, subjects select which statement best describes their health on that day from a possible 5 options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems). A unique EQ-5D health state is referred to by a 5-digit code allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the 5 dimensions. These data will be converted into a weighted health state index by applying

scores from EQ-5D value sets elicited from general population samples (the base case will be the United Kingdom valuation set, with other country value sets applied in scenario analyses). Where values sets are not available, the EQ-5D-5L to EQ-5D-3L crosswalk will be applied.

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

3.1.10 Health resource utilization

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

3.2 Safety variables

Safety and tolerability will be assessed in terms of AEs, deaths, laboratory data, vital signs (pulse, BP and weight), ECG, physical examinations, and WHO performance status. These will be collected for all subjects.

3.2.1 AEs

All AEs will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Event (CTCAE) (version 4.0.3 or later).

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 are as follows: ocular surface effects, dry eye/conjunctivitis, interstitial lung disease and acute interstitial pneumonitis, nail and nail bed conditions, skin disorders, upper GI tract inflammatory events, cardiotoxicity, and diarrhoea.

Any AE occurring before treatment with AZD9291/chemotherapy 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 AZD9291/chemotherapy) will be included in the AE summaries. Any events in this period that occur after a subject has received further therapy for cancer (following discontinuation of AZD9291) will be flagged in the data listings. Please refer to CSP Section 6.3.1

3.2.2 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 treatment 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, BP and weight), and/or ECG data will be performed for identification of OAEs.

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

3.2.3 dECG

Baseline is the pre-dose time point on the first dosing day (Day 1 Cycle 1). For each time point three dECG recordings should be taken within an approximate 5 minute period. The variables to be reported from the continuous dECG measurements are RR, PQ, QRS, and QT intervals. The QT interval will be corrected for RR (the duration of a heart beat) to obtain corrected (QTc) variables.

QTcF will be calculated using the formula:

$$QTcF=QT/RR_b$$

Where b=1/3.

To obtain a single value of QTcF, RR, PR, QRS and QT at each specified time point, the mean of the triplicate values at that time point will be used. If there are more than three values for a single time point, the mean of all the recorded values will be taken. For each patient, the change-from-baseline in a dECG variable at each time point will be calculated as the difference between the mean of the replicate value at each post-dose time point and the mean of pre-dose baseline replicate value. All ECG data will be obtained from the vendor, eRT, and there will be no site interpretation.

3.2.4 Duration of exposure

Exposure to AZD9291

Total exposure to AZD9291 will be time (days) from the first dose to the last dose:

> Total exposure = (last dose date where dose >0 mg - first dose date) + 1

Actual exposure to AZD9291 will be time (days) from first dose to the last dose, taking account of dose interruptions.

> Actual exposure = ((last dose date where dose >0 mg - first dose date) + 1) - total duration of dose interruption (i.e., number of days with dose = 0mg).

Exposure to chemotherapy

Duration of treatment on chemotherapy will be in terms of the number of cycles. A cycle will be counted if chemotherapy is started, even if the full dose is not delivered.

If a new cycle is delayed, the day when chemotherapy dosing occurs will be deemed to be day 1 of the next cycle.

3.2.5 Dose intensity

Relative dose intensity (RDI) and percentage intended dose (PID) will be defined as follows:

- RDI = 100% x d / D, where d is the actual cumulative dose delivered up to the earlier of progression (or a censoring event) or the actual last day of dosing, and D is the intended cumulative dose up to the earlier of progression (or a censoring event) or the actual last day of dosing plus the protocol-defined post-dose rest period. Protocol-defined post-dose rest period is 0 days for AZD9291.

- PID = 100% x d_1 / D_1, where d_1 is the actual cumulative dose delivered up to progression (or a censoring event), and D_1 is the intended cumulative dose up to progression (or a censoring event). D_1 is the total dose that would be delivered, if there were no modifications to dose or schedule.

3.3 PK variables

PK analysis of the plasma concentration data for AZD9291, AZ5104 and AZ7550 will be performed by Quantitative Clinical Pharmacology, AstraZeneca or delegate on behalf of Quantitative Clinical Pharmacology. Plasma concentration will be listed and summarized according to the time ranges.

The ratio of metabolite to AZD9291 will be calculated at the time-points. The plasma concentration data for AZD9291 and metabolites will also be analyzed using a population PK approach, which may include exploring the influence of covariates on PK, if the data allows.

The data collected in this study may also be combined with similar data from other studies and explored 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.

The concentration in CSF will be summarized and metabolite to AZD9291 ratio calculated. If the CSF was taken on a PK day then the CSF to pre-dose plasma concentration for AZD9291 will be calculated. If there are no CSF samples collected in the study, CSF data will not be presented or summarized.

4. ANALYSIS METHODS

4.1 General principles

Continuous data will be summarized using descriptive statistics (number of observations, mean, standard deviation, median, 25th and 75th percentiles, minimum, and maximum). Frequencies and percentages will be used for summarizing 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 treatment group and subject number.

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

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.

4.1.1 Baseline measurements and change from baseline variables

In general, for efficacy endpoints and PROs the last observed measurement prior to randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before the first dose of randomized treatment then this assessment will be used as baseline. For safety endpoints the last observation before 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 the first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

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

4.1.2 Multiple testing strategy

In order to provide strong control of the type I error rate, the primary endpoint of PFS and key secondary endpoints, ORR and OS will be tested in this sequential order. If any previous

analysis in the sequence is not statistically significant, the alpha cannot be transferred to subsequent analyses.

The analyses of primary and all secondary endpoints will all occur at the time of primary analysis of PFS with the exception of OS. Three analyses of OS are planned; two interim analyses and a final analysis. The data cut-off for the first analysis of overall survival will be conducted approximately 4 months after data cut-off for the primary analysis of PFS and the second analysis will be conducted when the OS data are approximately 50% mature (approximately 205 death events). A final analysis of overall survival will be performed when the OS data are approximately 70% mature (approximately 287 death events).

The analyses of PFS2, time to first subsequent therapy and time to second subsequent therapy will be performed at the time of the OS analyses since these endpoints are considered important to the interpretation of OS.

A 2-sided 5% alpha will be used in all testing, with the exception of overall survival endpoint. Since three 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 three 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 chemotherapy arm of 16 months and a median OS of 22 months on the AZD9291 arm, that 119 OS events were observed at the first analysis, the information fraction would be entered as 0.41 (119/287 events) for the first analysis and 0.71 (205/287 events) for the second analysis, since 205 and 287 events are expected at the second and final analysis, respectively. This would result in a significance level for the first analysis of 0.001 (2-sided), a significance level for the second analysis of 0.016 (2-sided) and a significance level for the final analysis of 0.045 (2-sided).

Note, any non-statistically significant analyses at the interims will not preclude further testing of OS.

For PRO symptoms, the overall type I error (5% 2 sided) will be controlled across the five primary PRO measures of cough, dyspnoea and pain as assessed by the EORTC QLQ-LC13 and fatigue and appetite loss as assessed by the EORTC QLQ-C30 using the Bonferroni-Holm procedure.

4.1.3 Study day definitions

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

that occur prior to randomization, study day is defined as (date of visit [event] - date of randomization). 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.

4.1.4 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 have the potential to be summarized. 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). 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 summarized, 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 subjects in either treatment group have data recorded at a particular visit.

4.1.5 Handling missing data

In general, other than for partial dates, missing data will not be imputed and will be treated as missing.

Imputation of partial dates

Initial diagnosis date:

- If year is missing (or completely missing), do not impute.
- If only day is missing, impute day as 15th of the month.
- If day and month are missing, impute as July 1st.

Concomitant medication start date

- 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 end date

- 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.
- If year and month are present and day is missing, impute day as last day of the month.

4.1.6 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 x 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 x 1.01 for analysis purposes. The original value will be listed.

4.1.7 Rounding rules for reported percentages

For percentages ≥10%:

- Values ≥X.5 or above round to X+1.

- Values >X but <X.5 round to X.

For percentages <10%:

- Values ≥X.Y5 or above round to X.Y+0.1.
- Values >X.Y but <X.Y5 round to X.Y.

4.2 Analysis methods

4.2.1 Subject disposition and data sets analyzed

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

- Randomized
- Treated
- Subjects ongoing study treatment at the data cut-off
- Included in each analysis set (FAS, Safety, and PK).

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

4.2.2 Protocol deviations

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

4.2.3 Demographic and other baseline characteristics

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

- Age (years).
- Weight (kg).
- Height (cm).
- Body mass index (kg/m^2) [calculated as ($weight/height^2$) where weight is in kg and height is in m].
- Nicotine consumption (cigarettes/cigars per day and pipe tobacco packs per week)

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

- Age group (years) (grouped as <50, ≥50-<65, ≥65)
- Sex
- Race
- Ethnic group
- EGFR mutation type (Exon 20 T790M, G719X, S768I, Exon 19 deletion, L858R, and other) (as reported prior to study screening)
- Overall disease classification (metastatic, locally advanced, both)
- Site of disease
- Baseline TL size (mean and categories: <40, 40-79, 80-119, ≥120mm)
- Histology
- Smoking status.
- WHO performance status (0/1)

The number of subjects with all types of EGFR mutations identified by the cobas EGFR central test (by biopsy taken after confirmation of disease progression on the most recent treatment regime) will be summarized (i.e., T790M only, T790M and exon 19 deletion, T790M and L858R, and T790M with all other mutations with >5 subjects). Examples of other EGFR mutations which may be detected are G719X, S768I, and Exon 20 insertion). A listing of all subjects showing all the EGFR mutations reported prior to study screening and the EGFR mutations identified by the cobas central test after confirmation of disease progression will be presented.

4.2.4 Medical history

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

All relevant surgical history will be listed and summarized similarly.

4.2.5 Anti-cancer medications and radiotherapy (prior, during, post-treatment with IP)

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

Concomitant medications/radiotherapy will be listed and summarized 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 4.1.5.

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

- Prior medications are those taken prior to screening 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.

EGFR-TKI therapies will be defined as approved EGFR-TKI therapies and will include gefitinib, erlotinib and afatinib only.

4.2.5.1 Anti-cancer therapy

All anti-cancer therapies will be summarized for the FAS. They will be summarized separately for prior, current, and post-withdrawal of IP anti-cancer therapies.

The number of anti-cancer therapy regimens and number of cycles will be summarized using descriptive statistics. The anti-cancer therapy agent, therapy class, best response, and treatment status will be summarized using frequencies and percentages. Any anti-cancer therapies will also be listed.

4.2.5.2 Radiotherapy

All radiotherapies will be summarized for the FAS. They will be summarized separately for prior, current, and post-withdrawal of IP radiotherapies.

The number of fraction doses and fraction dose (Gy) will be summarized using descriptive statistics. The radiotherapy site/region and treatment status will be summarized using frequencies and percentages.

Any radiotherapy will also be listed.

4.2.6 Exposure

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

- Summary of duration of exposure of AZD9291
- Summary of total number of cycles of chemotherapy received
- RDI and PID of AZD9291
- Summary of interruptions and reductions of AZD9291
- Summary of delays and reductions of chemotherapy

4.2.7 Brain metastases

The proportion of subjects who develop brain metastases during the trial will be summarized by treatment group.

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

4.2.8 Efficacy

All efficacy analyses will be performed on the FAS. Results of all statistical analyses will be presented using a 95% CI and 2-sided p-value.

The stratification factor for the efficacy analyses will be based on the values entered into the IVRS.

4.2.8.1 Primary outcome: PFS

PFS will be analyzed using a log rank test stratified by ethnicity (Asian, Non-Asian) 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 G et al 1991, Selke & Siegmund 1982):

$$HR = \exp(U/V)$$

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

Where $U = \Sigma_i(d_{1i} - e_{1i})$ is the log-rank test statistic (with d_{1i} and e_{1i} the observed and expected events in group 1) and $\sqrt{V}$ the standard deviation of the log-rank test statistic obtained from the LIFETEST procedure with a STRATA term for the stratification variable.

A Kaplan-Meier (KM) plot of PFS will be presented by treatment group.

Subgroup analysis

Subgroup analyses will be conducted by comparing PFS between treatments in the following groups:

- Ethnicity (Asian versus Non-Asian)

- Gender (Male versus Female)
- Age at screening (<65 versus ≥65)
- Mutation status prior to start of study (ex19 versus L858R)
- Duration of prior EGFR-TKI (<6 months, ≥6 months)
- Brain metastases at entry
- Smoking history

The HR (AZD9291: chemotherapy) and associated CI will be calculated from a Cox proportional hazards model (ties=Efron) that contains the treatment term, factor and treatment-by-factor interaction term. The treatment effect HRs for each treatment comparison along with their CIs will be obtained for each level of the subgroup from this single model. The HRs and 95% CIs will be presented on a forest plot including the HR and 95% CI from the overall population (using the primary analysis).

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 analyzed. In this case, only descriptive summaries will be provided.

The assumption of proportionality will be assessed. Proportionality will be tested firstly by examining the plots of complementary log-log (event times) versus log (time) and, if necessary, a time dependent covariate will be fitted to assess the extent to which this represents random variation.

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

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, all covariates, and all covariate-by-treatment interaction terms, with one that excludes the interaction terms and will be assessed at the 2-sided 10% significance level. If the fit of the model is not significantly improved as assessed by twice the change in log-likelihood then it will be concluded that overall the treatment effect is consistent across the subgroups.

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

the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

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

Sensitivity analysis

(a) Ascertainment bias

The possibility of bias in assessment and measurement of PFS by investigators will be assessed by comparing the hazard ratio derived from investigator review with the hazard ratio derived using the BICR assessment of disease progression by RECIST 1.1.

(b) Evaluation-time bias

In order to assess possible evaluation-time bias, which 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 1.1 assessment will be analyzed using a stratified log rank test, as described for the primary analysis of PFS. For subjects who die in the absence of progression, the date of death will be used to derive the PFS time used in the analysis.

(c) Attrition bias

Possible attrition bias will be assessed by:

(i) Repeating the primary PFS analysis, except that the actual PFS event times, rather than the censored times, of subjects who progressed or died in the absence of progression immediately following two, or more, non-evaluable tumor assessments will be included.

(ii) Subjects who take subsequent therapy prior to progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy.

(iii) A reversed KM plot of the time to censoring, where the censoring indicator of the primary PFS analysis is reversed, will be presented to assess the number of subjects being followed over time.

Post-hoc exploratory analysis of the use of a novel sample BICR methodology to assess bias in assessment of PFS by investigators

Recently, it has been suggested that, in order to assess the potential for bias amongst investigators, a BICR amongst only a sample of patients could be performed; if evidence of

bias is detected, according to a pre-defined threshold, the BICR is then assessed in all patients, otherwise it is concluded the sample was sufficient to rule-out meaningful levels of bias. A proposed sample BICR method has been submitted to a peer reviewed statistical journal for publication as of August 2014 and in order to demonstrate that this is a valid methodology for use in future trials, a post hoc exploratory analysis will be carried out using this methodology. The results of this exploratory analysis will be reported separately from the CSR. A random sample of subjects from the full study population will be selected and the results of the BICR review of these subjects will be used to rule out a concerning level of bias in the full study local evaluation (LE) estimate of the treatment effect (Hazard Ratio) for PFS. The hazard ratio ratio (HRR), the ratio of the hazard ratio (HR) for the treatment effect estimated from the sample BICR to the corresponding HR for the investigator assessments, will be used as the metric to assess bias. Lack of evidence of meaningful bias will be concluded if there is a low conditional probability that the HRR seen in the random sample would have been observed, if in fact, the point estimate in the full trial were unacceptably high (i.e., 1.25).

The subjects forming the sample BICR will be randomly selected from within each treatment group with separate sampling within subjects who have a LE progression event and from those censored for PFS by the LE. This is so that the BICR sample is representative of the study as a whole.

A sample BICR would be deemed to be acceptable if there was a low (conditional) probability (10%), given the sample, that the null hypothesis of concerning bias in the overall study (HRR $\geq$1.25) was really true. This protects the sensitivity of the procedure at 90% (i.e., correctly failing to reject the hypothesis of concerning bias when this is really true).

4.2.8.2 Secondary outcomes

ORR

The ORR will be compared between treatment using logistic regression models adjusting for the covariate ethnicity (Asian/non-Asian). The results of the analysis will be presented in terms of an odds ratio together with its associated 95% profile likelihood CI. The p-value will be based on a test statistic that is calculated as twice the change in log-likelihood resulting from the addition of a treatment factor to a model that contains the covariate defined above. The test statistics is chi-squared distributed with 1 degree of freedom.

A summary of best objective response will also be presented by treatment group.

DoR

In order to analyze the secondary outcome variable of DoR between arms the Expected Duration of Response (EDoR) will be derived for each treatment group (Ellis S 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 all randomized subjects.

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 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. 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. Additionally, descriptive data will be provided for the duration of response in responding subjects, including associated KM curves (without any formal comparison of or p-value attached).

DCR

DCR will be analyzed using a logistic regression with a covariate for ethnicity (Asian/non-Asian). 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. 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.

Tumor shrinkage

The effect of AZD9291 on best percentage change in tumor size will be estimated from an analysis of covariance (ANCOVA) model with a covariate for ethnicity (Asian/non-Asian). 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.

OS

OS data will be analyzed using a log rank test stratified by ethnicity (Asian, Non-Asian) 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 described in Section 4.2.8.1 (provided there are sufficient events available for a meaningful analysis [>20 deaths], if not descriptive summaries will be provided). For the interim analysis, see Section 5.

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

Change from baseline in the primary PRO symptom scores of dyspnea, cough, pain, fatigue and appetite loss (described in Section 3.1.7) will be analysed using a mixed model for repeated measures (MMRM) analysis. The analysis will be to compare the average treatment effect from the point of randomisation for the first 6 months unless there is excessive missing data (defined as >75% missing data). Following confirmed disease progression, patients on the chemotherapy arm are eligible to crossover to AZD9291 treatment, therefore only data up to 6 months is included in this analysis to limit any confounding caused by treatment

switching. However, if the time to first AZD9291 dose when approximately 50% of chemotherapy patients receive AZD9291 occurs prior to 6 months post-randomisation then only data until this earlier time-point will be used in the analysis.

It is acknowledged that patients will discontinue treatment at different timepoints during the study and that this is an important time with regards to symptoms and HRQoL data collection. To account for this, and in order to include the discontinuation and follow up visits, a generic visit variable will be derived for each subject in order that the average treatment effect can be analyzed using the above method. Each visit will be assigned a sequential number. The time from randomization to each of these will be derived in order to select only those visits occurring within the first 6 months of randomization.

As an example, say a patient X attends the first 4 scheduled visits of a 6-weekly schedule and then discontinues treatment, whilst patient Y discontinues treatment after the first scheduled visit, the first 5 generic visits would be as follows:

Generic visit	Days since randomization	
	Patient X	Patient Y
Baseline	Baseline	Baseline
1	42	42
2	84	50 (discontinuation)
3	126	84
4	168	126
5	189 (discontinuation)	168

Thus, the first 5 generic visits could be used in the analysis of deriving the average treatment effect.

The MMRM model will include patient, treatment, and visit) and treatment by visit interaction as explanatory variables, and the baseline symptom score as a covariate along with the baseline symptom 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 and p-value. 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.

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.

Time to PRO symptom deterioration

Time to PRO symptom deterioration will be analyzed using a log rank test stratified by ethnicity (Asian/non-Asian) for generation of the p-value and using the Breslow approach for handling ties. The HR and CI will be estimated using the same methods as for the primary endpoint of PFS, as described in Section 4.1.

PRO symptom improvement rate

Symptom Improvement Rate at each visit will be analyzed using a logistic regression with a covariate for ethnicity (Asian/non-Asian). 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. 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.

Subjects with EGFR T790M mutation status detected in baseline plasma sample (ctDNA)

The efficacy of AZD9291 versus chemotherapy will be evaluated in patients with baseline plasma T790M mutation status (positive, negative and unknown). For PFS, the HR and associated 95% CI will be calculated for each group. For ORR, the odds ratio and associated 95% CI will be presented for each group.

Note, plasma testing will not be performed on China patients due to sample export limitations, therefore the denominator for this analysis will exclude patients from China.

4.2.8.3 Exploratory outcomes

PRO-CTCAE

PRO-CTCAE data will be presented using summaries and descriptive statistics based on the FAS.

Healthcare resource use

Healthcare Resource Use data (i.e., the hospital admission details including type of attendance and primary symptom for admission) will be presented using frequencies and percentages the duration of admission/attendance and duration of ICU/HDU stay will be summarized using descriptive statistics, based on the FAS.

Post-progression outcomes

PFS2, TFST and TSST will be analyzed using the same method as the analysis of PFS.

Further analysis of OS

Additional analysis of OS adjusting for the impact of subjects randomized to chemotherapy, who subsequently receive AZD9291 may be completed if this treatment sequence occurs in a significant proportion of subjects. Methods such as Rank Preserving Structural Failure Time (RPSFT) (Robins et al 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 a review of the data and the plausibility of the underlying assumptions. Further detail will be provided in a Payer Analysis Plan.

All further analyses of OS will be reported separately from the CSR.

4.2.9 PK and CSF concentration data

All plasma concentrations of AZD9291, AZ5104 and AZ7550 will be listed.

Secondary outcomes include summaries of plasma concentrations of AZD9292 and plasma metabolite (AZ5104 and AZ7550).

Exploratory outcomes include summaries of CSF concentrations, CSF metabolite: AZD9291 ratio, and CFS: plasma ratio, if any CSF data is obtained. If no CSF data is obtained, then this exploratory outcome will not be listed or summarized.

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

- The geometric mean (gmean, calculated as exp [μ], where μ is the mean of the data on a logarithmic scale).
- Coefficient of variation (CV, calculated as $100 \sqrt{[\exp(s^2)-1]}$, where s is the standard deviation of the data on a logarithmic scale).
- Gmean ± standard deviation (calculated as exp[$\mu \pm s$]).
- Arithmetic mean calculated using untransformed data.
- Standard Deviation (SD) 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 ± SD, 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 ± SD, arithmetic mean and SD 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 ± SD and SD 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 fewer than 3 subjects, no summary statistics other than minimum, maximum and n will be presented.

The PK data for Cycle 1 Day 1 and Cycle 3 Day 1 will also be displayed graphically for the PK analysis set. Displays will include plasma concentration subject profiles versus time and gmean concentrations (± SD) both on the linear and on the log-scale versus time by cohort.

4.2.10 Safety and tolerability

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

4.2.10.1 AEs

All AEs, both in terms of MedDRA PT and CTCAE grade, will be listed and summarized descriptively by count (n) and percentage (%) for each treatment group. MedDRA will be used for coding. Any AE occurring before AZD9291/chemotherapy treatment (i.e., before Study 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. The 28 day follow-up period will be defined as 28 days following discontinuation of AZD9291/chemotherapy treatment.

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved), investigator’s assessment of severity and relationship to study drug. Frequencies and percentages of subjects 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 subjects by treatment) 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 study medication
- AEs with outcome of death
- AEs with outcome of death causally related to study medication
- All SAEs
- All SAEs causally related to study medication
- AEs leading to discontinuation of AZD9291/chemotherapy
- AEs leading to discontinuation of AZD9291/chemotherapy, causally related to AZD9291/chemotherapy
- OAEs
- OAEs causally related to AZD9291/chemotherapy

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

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

Summaries of the number and percentage of subjects with AEs leading to dose interruptions of AZD9291/chemotherapy will be presented by PT.

In addition, AEs with an outcome of death, SAEs, AEs leading to discontinuation of treatment, AEs causally related to AZD9291/chemotherapy and OAEs will be listed.

A separate summary will be produced that presents any events that occur prior to dosing or starting more than 28 days after discontinuing therapy.

4.2.10.2 AEs of Special Interest

Summary tables of AEs of special interest will be produced including the number (%) of patients experiencing any of the specified terms presented by maximum CTCAE grade, time to onset of first AE for each grouped term and each preferred term within it, time to onset of first CTCAE grade 3 or higher, fluctuations in CTCAE grade, duration of AEs of special interest and number (%) of patients treated for AEs of special interest.

Life table and prevalence plots for rash (rashes and acnes), diarrhoea and fatigue will be produced.

4.2.10.3 Deaths

A summary of deaths will be provided with the number and percentage of subjects, categorized 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
- Subjects with unknown reason for death.

A corresponding listing will also be produced.

4.2.10.4 Laboratory evaluations

All laboratory data recorded in the electronic case report form (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 reference ranges provided by the local or central laboratory. As applicable, values will be graded using CTCAE v4.0.

For clinical chemistry and hematology, shift tables of the number of CTCAE grade changes from baseline will be provided. Corresponding shift tables (“Negative”, “Trace”, “Positive”, “0”, “+”, “++”, “+++”) will be produced for urinalysis.

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

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

The following laboratory variables will be summarized:

Table 3 Laboratory safety variables

Clinical chemistry (Serum/Plasma)	Hematology (Blood)	Urinalysis (Urine)
Albumin	Hemoglobin	Glucose
ALT	Leukocyte	Protein
AST	Hematocrit	Blood
Alkaline phosphatase	Red blood cell (RBC) count	
Bilirubin, total	Absolute leukocyte differential count:	
Calcium, total	Neutrophils	
Creatinine	Lymphocytes	
Calculated creatinine clearance	Monocytes	
Glucose (fasting, on PK days only)[1]	Basophils	
LDH[2]	Eosinophils	
HbA1C	Platelet count	
Magnesium	Reticulocytes	
Potassium		
Sodium		
Urea nitrogen/Blood urea nitrogen (BUN)		

[1] Subjects will be required to fast (water only) for at least 8 hours prior to the collection of a fasting glucose sample required on PK days. Random glucose sample will be collected on non-PK days.
[2] LDH is an additional variable collected at Visit 1 only.

4.2.10.5 Vital signs (pulse and BP) and weight

All vital signs data will be listed. Absolute and percentage change from baseline for pulse, BP and weight will be summarized by treatment group and visit.

4.2.10.6 Physical examination

Abnormalities identified from physical examination will be listed. For each physical examination body system, the number and percentage of subjects with abnormalities at baseline and post-baseline will be summarized.

4.2.10.7 ECG

dECG

All dECG data received will be presented in data listings. QTc summaries will be presented for subjects in the safety analysis set.

The dECG parameters that will be summarized (absolute values and change from baseline) are: QTcF, RR, PR, QRS and QT.

The observed values and change from baseline in QTcF will be summarized by visit using summary statistics. Plots of mean observed QTc parameters and change from baseline in QTc parameters versus time will be presented. Shift plots of the maximum change from baseline versus the baseline value for QTcF, with reference lines for 450 ms, 480 ms, 500 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 summarized 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 subjects who meet the ECG outlier criteria at any assessment post-date of first dose will be summarized.

LVEF

LVEF outliers are defined as LVEF values following dosing that are

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

LVEF outliers will be highlighted in the data listings and summarized.

4.2.10.8 WHO performance status

WHO performance status will be listed and summarized by treatment and visit.

4.2.11 Crossover

Summaries for subjects randomized to the chemotherapy arm that crossover to receive treatment with AZD9291 will be provided and will include patient disposition, duration of

exposure to AZD9291, treatment interruptions and dose reductions, and adverse events summaries.

5. INTERIM ANALYSES

Three analyses of OS are planned:

- First analysis of OS will be conducted approximately 4 months after data cut-off for the primary analysis of PFS.
- Second analysis will be conducted when the OS data are approximately 50% mature (approximately 205 death events).
- A third analysis of OS will be performed when the OS data are approximately 70% mature (approximately 287 death events).

6. ADDITIONAL BRAIN METASTASES ANALYSIS

An additional exploratory analysis will be conducted to explore efficacy in 2 subgroups of patients.

a) The first set of analysis will be conducted to explore the efficacy in patients who undertook a brain scan in the screening/baseline period, had their scan sent for independent radiology review and were identified by that review as having non-measurable and/or measurable brain disease at baseline at the time of randomization.

b) The second set of analysis will be conducted to explore the efficacy in patients with and without a prior history of brain metastases and/or current brain lesion at baseline.

The analysis sets, definition of required variables and description of analysis will be pre-specified in a separate SAP.

7. CHANGES OF ANALYSIS FROM PROTOCOL

The China-cohort analyses will be performed when approximately 27 PFS events have occurred in approximately 50 Chinese patients so that the maturity of the PFS analysis is consistent with that of the global analysis. This update was not included in protocol amendment 3 in error.

The definitions of ORR and DoR have been amended so that visit responses of CR or PR do not need to be confirmed. This is in line with New response evaluation criteria in solid

tumours: Revised RECIST guideline (version 1.1) that, in randomised Phase III studies where progression is the primary endpoint, confirmation of response is not required.

The PK analysis set includes subjects in the FAS who have at least one measurable PK concentration and who have the relevant date, time and dosing data for this sample. The definition that the dosing data must include the dosing data for the 2 days prior to the sample day as well as the sample day has been changed to include the dosing data 1 calendar day prior to the sample day as there were very few samples for the 2 days prior to the sample day.

8. REFERENCES

Berry G et al 1991
Berry G, Kitchin RM, Mock PA. A comparison of two simple hazard ratio estimators based on the logrank test. Stat Med. 1991 May;10(5):749-55.

Ellis S et al 2008
Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials. Contemp Clin Trials. 2008 Jul;29(4):456-65.

Gail and Simon 1985
Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics. 1985 Jun;41(2):361-72.

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

Selke & Siegmund 1982
Selke T, Siegmund D. Sequential analysis of the proportional hazards model. Biometrika. 1983;70(2): 315-26.

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

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

9. APPENDIX

N/A