Statistical Analysis Plan	
Study Code	D1532C00079
Edition Number	3.0
Date	

A Phase III, Double-Blind, Randomised, Placebo-Controlled Study to Assess the Efficacy and Safety of Selumetinib (AZD6244; ARRY-142886) (Hyd-Sulfate) in Combination with Docetaxel, in Patients receiving second line treatment for *KRAS* Mutation-Positive Locally Advanced or Metastatic Non Small Cell Lung Cancer (Stage IIIB – IV) (SELECT-1)

A Phase III, Double-Blind, Randomised, Placebo-Controlled Study to Assess the Efficacy and Safety of Selumetinib (AZD6244; ARRY-142886) (Hyd-Sulfate) in Combination with Docetaxel, in Patients receiving second line treatment for *KRAS* Mutation-Positive Locally Advanced or Metastatic Non Small Cell Lung Cancer (Stage IIIB – IV) (SELECT-1)

A Phase III, Double-Blind, Randomised, Placebo-Controlled Study to Assess the Efficacy and Safety of Selumetinib (AZD6244; ARRY-142886) (Hyd-Sulfate) in Combination with Docetaxel, in Patients receiving second line treatment for *KRAS* Mutation-Positive Locally Advanced or Metastatic Non Small Cell Lung Cancer (Stage IIIB – IV) (SELECT-1)

TABLE OF CONTENTS PAGE

TITLE PAGE 1
SIGNATURE OF STUDY STATISTICIAN 2
SIGNATURE OF GLOBAL PRODUCT STATISTICIAN 3
TABLE OF CONTENTS 4
LIST OF ABBREVIATIONS 8
AMENDMENT HISTORY 11
1. STUDY DETAILS 12
1.1 Study objectives 12
1.1.1 Primary objective 12
1.1.2 Secondary objective 12
1.1.3 Exploratory objective 13
1.2 Study design 14
1.3 Number of subjects 16
2. ANALYSIS SETS 17
2.1 Definition of analysis sets 17
2.1.1 Full Analysis Set 17
2.1.2 Safety Analysis Set 17
2.1.3 PK Analysis Set 17
2.2 Important protocol deviations 19
3. PRIMARY AND SECONDARY VARIABLES 20
3.1 General principles 20
3.1.1 Study day 21
3.1.2 Visit windows 21
3.1.3 Baseline 22
3.1.4 Handling of missing data 22
3.1.5 *KRAS* mutation status 22
3.2 Derivation of RECIST 1.1 visit responses 23
3.2.1 Target lesions (TLs) 23
3.2.2 Non-target lesions (NTLs) and new lesions 27
3.2.3 Overall visit response 28
3.2.4 Blinded Independent Central Review (BICR) Assessment 28
3.2.5 Best objective response (BOR) 29
3.3 Primary efficacy outcome variable 31

3.3.1 Progression Free Survival (PFS) 31
3.4 Secondary efficacy outcome variables 31
3.4.1 Overall survival (OS) 31
3.4.2 Objective response rate (ORR) 32
3.4.3 Duration of response (DOR) 32
3.5 Safety and tolerability outcome variables 33
3.5.1 Adverse event 33
3.5.2 Laboratory 34
3.5.3 Vital signs 35
3.5.4 Electrocardiograms (ECG) 35
3.5.5 Echocardiogram 35
3.5.6 Physical examination 35
3.5.7 Ophthalmic assessments 36
3.5.8 Duration of exposure 36
3.5.9 Dose intensity 37
3.6 Pharmacokinetic variables 37
3.7 Symptom endpoints and Health-related quality of life (HRQoL) 38
3.7.1 Lung Cancer Symptom Scale (LCSS) 38
3.7.1.1 Time to symptom progression (TSP) 40
3.7.1.2 Symptom Improvement rate 41
3.7.2 Short-Form Health Survey (SF-36v2) 41
3.7.2.1 Health domain scales and Physical and mental component summary scores 42
3.7.3 PRO compliance rates 43
3.8 Exploratory variables 43
3.8.1.1 Health state utility 43
3.8.2 Healthcare Resource Use 44
4. ANALYSIS METHODS 44
4.1 General principles 44
4.2 Analysis methods 45
4.2.1 Multiple testing strategy 46
4.2.2 Progression Free Survival 47
4.2.2.1 Sensitivity Analyses for Primary Endpoint (PFS) 50
4.2.2.2 Subgroup analyses 52
4.2.2.3 Secondary analysis 53
4.2.2.4 Consistency of treatment effect between subgroups 53
4.2.3 Overall Survival (OS) 54
4.2.4 Objective response rate (ORR) 55
4.2.4.1 Change in tumour size 55
4.2.5 Duration of response 57
4.2.6 RECIST 1.1 data by BICR 57
4.2.7 Patient reported outcomes (PRO) 58
4.2.7.1 Lung cancer symptom scale (LCSS) 58

4.2.7.2 Short-Form Health Survey (SF-36v2) 59
4.2.8 Safety 59
4.2.8.1 Adverse events 59
4.2.8.2 Laboratory assessments 62
4.2.8.3 Neutropenia and infection 63
4.2.8.4 Infection diagnostic investigations 64
4.2.8.5 Vital signs 64
4.2.8.6 ECG 65
4.2.8.7 Echocardiogram 65
4.2.8.8 Physical examination 66
4.2.8.9 Other safety data 66
4.2.9 WHO performance status 66
4.2.10 PK data 66
4.2.11 Demographic data and baseline data 66
4.2.12 Exposure 67
5. INDEPENDENT DATA MONITORING COMMITTEE (IDMC) 68
6. REFERENCES 68
7. APPENDIX 71

LIST OF TABLES

Table 1 Summary of Outcome Variables and Analysis Populations....18
Table 2 Classification of cobas® *KRAS* Mutation Test results....22
Table 3 TL visit responses....23
Table 4 Scaling-up TL visit response....26
Table 5 Overall visit response....28
Table 6 MCID for average, ASBI and individual LCSS score: Visit score....39
Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses....45
Table 8 Laboratory parameters - Minimum/Maximum of interest....71
Table 9 AstraZeneca project-specific normal range for laboratory parameters....74
Table 10 CTCAE grades for laboratory parameters....75
Table 11 Adverse Events of Special Interest and grouped terms....77

LIST OF FIGURES

Figure 1 Study flow chart....16
Figure 2 Flow diagram for determining best objective response....30
Figure 3 Multiple testing procedures for controlling type-1 error rate....47

LIST OF APPENDICES

Appendix A Laboratory parameters – Minimum/Maximum of interest....71
Appendix B BICR Audit Plan....72
Appendix C AstraZeneca project-specific normal range for laboratory parameters....74
Appendix D NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 grades for laboratory parameters....75
Appendix E Adverse Events of Special Interest (AESIs) and Grouped Terms....77

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
AESI	Adverse events of special interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ASBI	Average Symptom Burden Index (of the Lung Cancer Symptom Scale)
AST	Aspartate Aminotransferase
AUC	Area under plasma concentration / time curve
AZ	AstraZeneca
bd	Twice daily
BICR	Blinded Independent Central Review
BOR	Best Objective Response
BP	Blood pressure
cfDNA	Circulating free tumour deoxyribonucleic acid
CI	Confidence Interval
Cmax	Maximum plasma concentration
CR	Complete response
CRF	Case Report Form (electronic/paper)
CSR	Clinical Study Report
CT	Computer Tomography
CTCAE	Common Terminology Criteria for Adverse Event
DBL	Database Lock
DCO	Data cut off
DNA	Deoxyribonucleic acid
DoR	Duration of Response
ECG	Electrocardiogram
FAS	Full Analysis Set
FSI	First Subject In
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
IDMC	Independent Data Monitoring Committee

Abbreviation or special term	Explanation
ITT	Intention to Treat
iv	Intravenous
IVRS	Interactive Voice Response System
KM	Kaplan-Meier
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
KRAS mutation positive	Mutations in codon 12/13 and/or 61 of KRAS gene have been detected
LCSS	Lung Cancer Symptom Scale
LE	Local evaluation
LLOQ	Lower Limit of Quantification
LVEF	Left Ventricular Ejection Fraction
MCID	Minimum Clinically Important Difference
MTP	Multiple Testing Procedure
MRI	Magnetic resonance imaging
MUGA	Multi Gated Acquisition Scan
NSCLC	Non-Small Cell Lung Cancer
NTL	Non-target Lesion
OAE	Other Significant Adverse Event
od	Once daily
ORR	Objective Response Rate
OS	Overall Survival
P	Probability
PD	Pharmacodynamic
PFS	Progression Free Survival
PK	Pharmacokinetic
PRO	Patient-Reported Outcome
QTcF	QT corrected using Fridericia's formula
RECIST 1.1	Response Evaluation Criteria in Solid Tumours version 1.1
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SF-36v2	36 Item Short-Form Health Survey Version 2
SD	Stable Disease

Abbreviation or special term	Explanation
TL	Target Lesion
TSP	Time to Symptom Progression
ULN	Upper Limit of Normal
VAS	Visual Analogue Scale
WHO	World Health Organisation

AMENDMENT HISTORY

Date	Brief description of change
	Updates to first edition of the SAP including: • Change in sample size per the CSP amendment • Removal of interim analysis • Addition of hospitalisation exploratory analyses • Modifications to present final agreed list of important protocol deviations • Addition of stratification inconsistency listing • Safety study day definition modification • Baseline definition modification • BICR process modification • AEs of special interest clarification • Clarification and addition of statistical models in proportionality assumption section • Changes to ascertainment bias analyses • Addition of more subgroups to subgroup analyses • Addition of change in tumour size section • Addition of rules for imputing a partial death date • Addition of SAEs in screening failure patients related to biopsy listing • Removal of AE event rate presentation • Addition of WHO performance status shift table • Addition of BICR, AZ project-specific normal ranges, CTCAE grades and adverse events of special interest appendices • Removal of SF-36 derivation of T-scores • Changes to *KRAS* mutation status section • Minor corrections and clarifications
	Updates to second edition of the SAP including: • Additional detail added to the important protocol deviation definitions • Clarification of number of days used to define months and years • Additional detail added for calculation of PID and RDI • Further details on the assessment of proportionality and analyses relating to this are included • Further detail on the 2 methods for analysing BICR data is included • Categorical analyses based on improvement/worsening of SF36 removed • Correction to the Hy's law criteria

1. STUDY DETAILS

1.1 Study objectives

1.1.1 Primary objective

Primary objective	Outcome variables
To assess the efficacy in terms of PFS of selumetinib in combination with docetaxel compared to placebo in combination with docetaxel.	• Progression Free Survival (PFS) using investigator site assessments according to Response Evaluation Criteria in Solid Tumours (RECIST) 1.1

1.1.2 Secondary objective

Secondary objectives	Outcome variables
To assess the efficacy of selumetinib in combination with docetaxel compared with placebo in combination with docetaxel.	• Overall survival (OS) • Objective Response Rate (ORR) using investigator site assessments according to RECIST 1.1 • Duration of Response (DoR) using investigator site assessments according to RECIST 1.1
To assess the efficacy of selumetinib in combination with docetaxel compared with placebo in combination with docetaxel on NSCLC symptoms.	Average Symptom Burden Index (ASBI) of the six symptoms (appetite, fatigue, coughing, shortness of breath, blood in sputum and pain) in the LCSS will be used to assess: • Time to symptom progression • Symptom improvement score
To assess the safety and tolerability profile of selumetinib in combination with docetaxel compared with placebo in combination with docetaxel	• Adverse Events (AEs) • Clinical chemistry, haematology and urinalysis • Vital signs • Electrocardiogram (ECG) • Echocardiogram (ECHO)/ Multi Gated Acquisition Scan (MUGA) • Ophthalmological examination

Secondary objectives	Outcome variables
To investigate the pharmacokinetics (PK) of selumetinib and N-desmethyl selumetinib when administered in combination with docetaxel (other selumetinib metabolites e.g. selumetinib amide, may also be assessed)	Where the data allow, derived PK parameters for selumetinib and N-desmethyl selumetinib will be produced which may include, but are not restricted to, Maximum plasma concentration (Cmax) and Area under plasma concentration / time curve (AUC)

1.1.3 Exploratory objective

Exploratory objectives	Outcome variables
To describe the impact of treatment (selumetinib in combination with docetaxel and with placebo in combination with docetaxel) and disease state on symptom distress and interference with activity levels as measured by the Lung Cancer Symptom Scale (LCSS) and Health-Related Quality of Life (HRQoL) as measured by SF-36v2	Changes within each of the two treatment groups in the individual items of the LCSS • Symptom Distress • Interference with activity levels SF-36v2 will be used to describe changes in 8 domain scores: • Physical functioning • Role limitations due to physical health problems • Bodily pain • Social functioning • General mental health • Role limitations due to emotional problems • Vitality, energy or fatigue • General health And the two component summary scores over time: • Physical component summary • Mental component summary
To investigate the relationship between selumetinib and/or N-desmethyl selumetinib plasma concentrations/exposure and clinical outcomes, efficacy, AEs and/or safety parameters if deemed appropriate*	Output from both graphical and/or appropriate Pharmacokinetic (PK)/ Pharmacodynamic (PD) modelling techniques.

Exploratory objectives	Outcome variables
To investigate the use of Circulating free tumour deoxyribonucleic acid (cfDNA) derived from plasma, for the analysis of KRAS mutation status at screening and treatment discontinuation*	*KRAS* mutation status of plasma derived DNA from samples collected at screening and treatment discontinuation
To explore the influence of *KRAS* mutation subtypes on response to treatment *	*KRAS* mutation subtype(s)
To collect and store DNA, derived from a blood sample, for future exploratory research into genes/genetic factors that may influence response e.g. distribution, safety, tolerability and efficacy of selumetinib and/or agents used in combination and/or as comparators (optional)*	Host genetic polymorphisms
To explore potential biomarkers in residual biological samples (e.g. tumour and/or plasma) which may influence development of cancer (and associated clinical characteristics) and/or response*	Biomarkers of response and/or development of cancer
To investigate the health economic impact of treatment and the disease on hospital related recourse use and health state utility.	Exploratory variables include number, type and reason of hospitalisations and hospital attendances, procedures undertaken and hospital length of stay. Health state utility derived from the HRQOL instruments, the SF-36 v2.

*: These exploratory objectives will be reported separately from the clinical study report (CSR) and the details of these analyses will not be specified in this statistical analysis plan (SAP)

1.2 Study design

This is a phase III, double-blind, randomised, placebo-controlled study assessing the efficacy and safety of selumetinib 75 mg bd in combination with docetaxel 75 mg/m^2 and placebo in combination with docetaxel 75 mg/m^2, in patients receiving second line treatment for *KRAS* mutation-positive locally advanced or metastatic (Stage IIIB – IV) NSCLC.

Patients will be enrolled on the basis of their NSCLC treatment status. The *KRAS* mutation status of the patient's tumour must be determined prospectively by a central laboratory using the cobas® *KRAS* Mutation Test under IDE G130187. AstraZeneca has partnered with to develop the *KRAS* mutation test as a tissue-based companion diagnostic for selumetinib. The term *KRAS* mutation positive is used to refer to any sample where mutations in codons 12/13 or 61 have been detected. Patients whose tumour sample harbour a *KRAS* mutation and fulfil all eligibility criteria will be randomised in a ratio of 1:1 to receive selumetinib 75 mg bd in combination with docetaxel 75 mg/m^2 or placebo in combination with docetaxel 75 mg/m^2. All patients will also receive 6 mg

pegylated G-CSF at least 24 hours after the administration of every docetaxel dose and not within 14 days prior to the next docetaxel administration.

Patients will be stratified at randomisation based on their World Health Organisation (WHO) Performance Status (1/0) and tumour histology (squamous/non-squamous). Thus, there will be four strata:

- WHO performance status = 1 and Histology = squamous
- WHO performance status = 0 and Histology = squamous
- WHO performance status = 1 and Histology = non-squamous
- WHO performance status = 0 and Histology = non-squamous

If the Investigator deems it is appropriate, patients may be given the option to consent for *KRAS* mutation status screening (at the designated central laboratory) prior to consenting to the main study. In this instance, archival tumour material should be provided for this assessment. Only data required for the *KRAS* mutation screening will be collected at this time such as demographic data, tumour status and prior cancer treatment. AE/SAE data collection is not required prior to main consent. If *KRAS* mutation positive status is confirmed, the patient should be given the option to consent to the main study and *KRAS* mutation screening will not need to be repeated during visit 1.

Patients will be seen and assessments performed as outlined in the study plan (see Table 3 (Study Plan) of the Clinical Study Protocol) until objective disease progression or until meeting a criterion for discontinuation from study treatment or from the study.

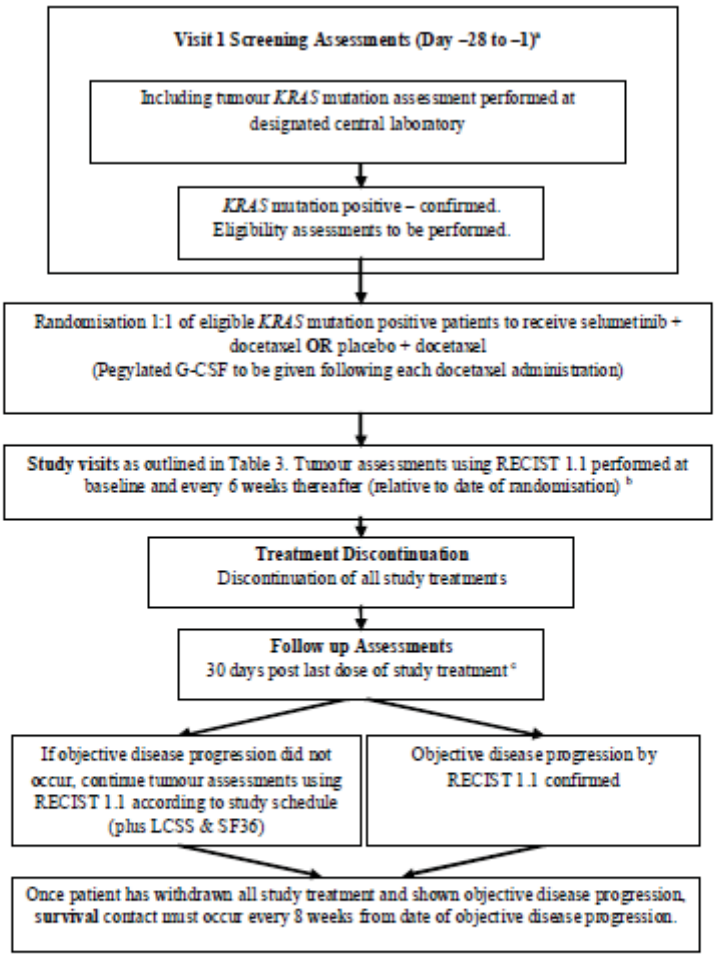
Patients may continue to receive study treatment (docetaxel, selumetinib/placebo), after objective disease progression if, in the opinion of the investigator, they are continuing to derive clinical benefit, in the absence of significant toxicity and it does not contravene local practice, after consultation with AZ. Patients remaining on study treatment beyond progression will continue to be seen as outlined in CSP study plan and will have all required assessments performed excluding RECIST 1.1 assessments and PRO questionnaires (last scheduled PRO questionnaire is to be completed approximately 30 days after progression).

Once a patient has had objective disease progression recorded and discontinued all study treatment, they are to be followed up for survival status every 8 weeks until death, withdrawal of consent or the end of the study, whichever occurs first (Figure 1).

An Independent Data Monitoring Committee (IDMC) will be established in order to assess the progress of the clinical trial, including reviewing the safety data. See Section 5 for further details.

The study flow chart is shown in Figure 1.

Figure 1 Study flow chart



[a] Screening assessments can be performed in a stepwise process or in parallel to *KRAS* mutation assessment. Note: If the patient provides consent for *KRAS* mutation status screening (prior to the main study consent) *KRAS* mutation status will be tested by the designated central laboratory prior to visit 1.
[b] Up until the time of data cut off (DCO) for the analysis of PFS, patients must be followed until evidence of RECIST 1.1 defined progression (regardless of reason for treatment discontinuation)
[c] Last dose of selumetinib/placebo or docetaxel.

1.3 Number of subjects

PFS is the primary endpoint for this study. However, the study has been sized to characterise the OS benefit of selumetinib 75mg bd in combination with docetaxel 75mg/m^2.

510 *KRAS* mutation positive tumour patients will be randomised (1:1) between the two treatment arms to obtain 332 death events (65% maturity). If the true OS hazard ratio (HR) for the comparison of selumetinib in combination with docetaxel vs placebo in combination with docetaxel is 0.72 (likely to correspond to a 38% prolongation of OS), the study has at least 80% power to demonstrate a statistically significant difference for OS, assuming a 2% 1-sided significance level (see Section 4.2.1 for details of the multiple testing strategy and control of the Type I error).

An OS HR of 0.72 corresponds to an approximate 2-month improvement in median OS over an estimate of 5.2 months (estimated from D1532C00016) for placebo in combination with docetaxel, assuming proportional hazards and exponential data distribution. A 2-month improvement in median OS is regarded as clinically meaningful. The smallest treatment difference that would be statistically significant at the 1- sided 2% level is an OS HR of 0.80 (0.796 if exactly 332 OS events).

With this number of patients and if the true PFS HR is 0.58, the study will provide over 90% power to demonstrate a statistically significant difference for PFS assuming a 2.5% 1-sided Type I error. This HR corresponds to the estimated treatment effect observed in the Phase II study D1532C00016, based on 71 events and a difference in medians of approximately 3 months.

2. ANALYSIS SETS

2.1 Definition of analysis sets

There are three analysis sets defined in this study which are as follows:

2.1.1 Full Analysis Set

Efficacy data will be summarised using the Full Analysis Set (FAS), following the principle of intention-to-treat (ITT). The FAS will include all randomised patients and will compare the treatment groups on the basis of randomised treatment, regardless of treatment actually received.

2.1.2 Safety Analysis Set

All patients who received at least one dose of randomised investigational product (selumetinib/placebo) will be included in the safety population.

For all safety endpoints, erroneously treated patients (e.g., those randomised to treatment A but actually given treatment B) will be accounted for in the actual treatment group. Treatment received is based on the initial dose of study treatment received, even though patients may have had subsequent dose reductions.

Note, any patient who received at least one dose of selumetinib will be included in the selumetinib group, even if the patient was planned to receive placebo.

Summaries of safety and tolerability data will be produced based on the safety analysis set.

2.1.3 PK Analysis Set

PK analysis set will include all patients who receive study treatment as per protocol and do not violate or deviate from the protocol in ways that would significantly affect the PK analyses.

The population will be defined by the Study Team Physician, Pharmacokineticist and Statistician prior to any analyses being performed. PK data will be analysed according to treatment received.

Table 1 Summary of Outcome Variables and Analysis Populations

Outcome Variable	Populations
Study Population Data	
Demography characteristics (e.g. age, sex etc.)	FAS
Baseline characteristics and disease characteristics (e.g. primary tumour location, nicotine use, WHO performance status, *KRAS* mutation subtype, histology etc.)	FAS
Analysis populations	FAS
Important deviations	FAS
Medical/Surgical history	FAS
Previous anti-cancer therapy	FAS
Concomitant medications/procedures	Safety
Subsequent anti-cancer therapy	FAS
Efficacy Data	
PFS	FAS
OS, ORR, DoR	FAS
Pharmacokinetic Data	
Pharmacokinetic data	PK
Patient reported outcomes (PRO)	
Symptom improvement rate[a]	FAS
Time to symptom progression[a]	FAS
Health-related quality of life (LCSS or SF-36v2)	FAS
Safety Data	
Exposure	Safety
Adverse Events	Safety
Lab measurements	Safety
WHO performance status	FAS
ECHO/MUGA	Safety
Vital Signs	Safety

[a] There will be difference in numbers of patients evaluable for each endpoint within FAS.

2.2 Important protocol deviations

The important protocol deviations will be listed and summarised by randomised treatment group.

None of the deviations will lead to any patient 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. Some deviations may be important for safety point of view but these may not affect the interpretation of the study results.

Eligibility criteria deviations are deviations from the protocol inclusion and exclusion criteria. Post-entry deviations are deviations from the protocol that occurred after the patient was randomised to the study.

The study team physician and statistician will identify, adjudicate and classify the important deviations from monitor reports and programmatic checks prior to database lock/un-blinding.

Examples of important deviations will include:

- Patients who deviate from key entry criteria per the CSP. These are inclusion criteria 3, 4, 6, 7, and 8, and exclusion criteria 2, 3, 4, 6, and 16 (PK only).
- Patients with baseline RECIST 1.1 assessment more than 28 days prior to randomisation.
- No baseline RECIST 1.1 assessment on or before date of first dose.
- Patients who discontinue study treatment (selumetinib/placebo and docetaxel) for reasons other than objective disease progression, have not withdrawn main consent and do not continue to have RECIST assessments. This only applies to patients who are still alive more than 14 weeks after their last RECIST assessment.
- Patients randomised but not dosed (either selumetinib/placebo or docetaxel or both).
- Received incorrect investigational dose of selumetinib/placebo (defined as starting dose not equal to 75mg bid or a dosing interruption of >28 days without treatment discontinuation or any other non-protocol defined dose level taken for a period that could potentially impact efficacy. The latter will be based on clinical and statistical review).
- Received incorrect investigational treatment.
- Received prohibited concomitant medication (limited to anti-cancer agents) at the same time as selumetinib/placebo or docetaxel or both.

- Incorrectly un-blinded patients.

The specific deviations to be checked are:

A misrandomisation is when a patient is not randomised or treated according to the randomisation schedule. It is envisaged that there will be 2 sub categories of this:

(i) Patients 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 patients may stop taking drug.

(ii) The patient receives a treatment pack with a different code to their randomisation code. However, the actual treatment may still match the randomised treatment. For example, a patient is given randomisation code 0001, which according to the randomisation schedule is placebo. However, at the randomisation visit they are given treatment pack 0003, but this still contains placebo.

If any patients have any other protocol deviations, which are considered important by the study team, these will also be described.

Due to the nature of the patient populations and the study design, particularly the continuous chronic use of selumetinib/placebo, minor protocol deviations such as occasional forgotten doses are expected and will not be discussed in the Clinical Study Report (CSR).

Separately, the inconsistencies in stratification values (WHO Performance Status / tumour histology) between the Interactive Voice Response System (IVRS) and the RAVE database will be listed.

3. PRIMARY AND SECONDARY VARIABLES

3.1 General principles

Unless otherwise specified, data summaries and listings will be presented by the initial treatment group a patient was assigned to, i.e., initial randomised treatment or initial treatment received; even though patients may have had subsequent dose reductions. However, some listings such as AEs listings will display the actual dose the patient received at onset of an AE.

Unless otherwise stated, listings will include all data, including data recorded either before or after the study treatment period, and data collected at unscheduled visits.

When calculations are performed based on months one month will be considered to be 30.4375 days. When calculations are performed based on years one year will be considered to be 365 days.

3.1.1 Study day

Whenever data is summarised over time, study day will be calculated based on the actual assessment date. Efficacy data, HRQoL data and WHO performance status will be summarised in relation to date of randomisation, whereas safety data will be summarised in relation to date of first dose of any study treatment (selumetinib/placebo or docetaxel).

If actual assessment date is prior to randomisation/first dose then study day will be:

Study day = Actual assessment date - Randomisation/first dose date.

If actual assessment date is on or after randomisation/first dose then study day will be:

Study day = Actual assessment date - Randomisation/first dose date + 1.

3.1.2 Visit windows

For summaries of vital signs, laboratory data and echocardiogram scans, HRQoL, Patient reported outcomes etc., assessments will be assigned to calculated visit windows (using study day).

The time windows should be exhaustive so that data recorded at any timepoint has the potential to be summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit. For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit based summary, when deriving a patient 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 patient; they should also highlight the value for that patient that was used in the summary table, wherever feasible.

For visit based summaries:

- If there is more than one value per patient within a time window then the closest to the planned study day value should be summarised, or the earlier in the event the values are equidistant from the planned study day. The visit will be missing if no assessment was reported within the specified visit window around the planned study day.

- To prevent very large tables or plots being produced that contain many cells with meaningless data, summary statistics will be presented where at least 10 patients in either treatment group have data recorded at a particular visit.

3.1.3 Baseline

Baseline will be the last non-missing assessment of the variable under consideration prior to the intake of the first dose of any study treatment (selumetinib/placebo or docetaxel), for safety variables.

For variables summarised using the full analysis set, baseline is defined as the last non-missing assessment prior to randomisation. If a patient does not have an assessment prior to randomisation, the post-randomisation assessment would be acceptable as the baseline assessment if it were prior to the intake of the first dose of study treatment.

3.1.4 Handling of missing data

Missing Pharmacokinetic and Safety data will generally not be imputed.

However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification (LLOQ)) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings.

See sections 3.2 to 3.4 for missing data rules for efficacy endpoints.

3.1.5 *KRAS* mutation status

Patients with *KRAS* mutation positive tumours are prospectively selected using the cobas® *KRAS* Mutation Test. Each patient must have a *KRAS* mutation positive tumour status determined by the designated testing laboratory (approved lab) in order to receive treatment in this study.

Results for the cobas® *KRAS* Mutation Test can be reported as "Mutation Detected" or "No Mutation Detected" for the overall *KRAS* mutation status.

The cobas® *KRAS* Mutation Test also reports individual mutation status for patients who have a "Mutation Detected" result. These results are reported as codon 12/13 or codon 61. The patients with mutations detected in codon 12/13 as well as in codon 61 will be grouped into the subtype codon 12/13. The patients with a mutation detected in codon 61 alone will be grouped into the subtype codon 61. Thus patients can be grouped into *KRAS* mutation subtypes as shown in Table 2 below.

Table 2 Classification of cobas® *KRAS* Mutation Test results

cobas® *KRAS* Mutation Test result	**Mutation result**	***KRAS* Mutation Subtype**
Mutation Detected	Codon 12/13	Codon 12/13
Mutation Detected	Codon 12/13, codon 61	Codon 12/13
Mutation Detected	Codon 61	Codon 61

3.2 Derivation of RECIST 1.1 visit responses

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

Baseline radiological tumour assessments are to be performed no more than 28 days before the start of randomised treatment. Tumour assessments are then performed every 6 weeks（± 1 week）following randomization until disease progression according to the RECIST 1.1 criteria.

At each visit, an overall visit response will be determined programmatically - using the information from target lesions (TL), non-target lesions (NTL) and new lesions. RECIST 1.1 outcomes will be calculated using a computer program.

3.2.1 Target lesions (TLs)

Measurable disease is defined as having at least one measurable lesion (not previously irradiated) which is ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with computed tomography (CT) or magnetic resonance imaging (MRI). A patient can have a maximum of 5 measurable lesions recorded at baseline and these are referred to as target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest to randomisation will be used to define the baseline sum of TLs.

Measurable disease is one of the entry criteria for the study. However, if a patient with non-measurable disease is enrolled in the study, the evaluation of overall visit responses will be based on overall non-target lesion (NTL) assessment and the absence/presence of new lesion(s) (see section 3.2.2).

Table 3 provides details for TL visit responses.

Table 3 TL visit responses

Visit responses	Description
Complete Response (CR)	Disappearance of all target lesions. Any pathological lymph nodes selected as target lesions must have a reduction in short axis to <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum of diameters as long as criteria for PD are not met.

Table 3 TL visit responses

Visit responses	Description
Progressive disease (PD)	A ≥ 20% increase in the sum of diameters of target lesions and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
Not Evaluable (NE)	Only relevant in certain situations (i.e. if any of the target lesions were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a target lesion response
Not applicable (NA)	No target lesions are recorded at baseline

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to 1 d.p. before assigning a target lesion response. For example 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

Missing TL data

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

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

Lymph nodes

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

TL Visit responses subsequent to CR

A CR response can only be followed by CR, PD or NE. If a CR has occurred then the following rules at the subsequent visits must be applied:

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

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

TL too big to measure

If a target lesion 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 target lesion 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 target lesion becomes too small to measure a value of 5mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and entered a smaller value that can be reliably measured. If a target lesion 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 / embolisation), should be handled in the following way and once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours:

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

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and scale 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 would be assigned a TL response of PD.

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

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

Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of target lesion measurements are treated as missing (because of intervention) then target lesion response will be NE, unless the sum of diameters of non-missing target lesion 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 target lesions has increased by 5mm from nadir).

If ≤ 1/3 of the target lesion measurements are treated as 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.

Table 4 provides an example of scaling-up.

Table 4 Scaling-up TL visit response

Lesion	Longest diameter at nadir visit	Longest diameter at follow-up visit
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 has intervention thus treated as 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 baseline visit is 26.8 cm.

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

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

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

Change in method of assessment of TLs

CT and MRI are the only methods of assessment that can be used within a trial and clinical examination will not be used for assessment of TL. If a change in method of assessment occurs between CT and MRI this will be considered acceptable and no adjustment within the programming is needed.

3.2.2 Non-target lesions (NTLs) and new lesions

Non-target lesion response will be derived based on the Investigator's overall assessment of NTLs as follows:

Progressive disease:	Unequivocal progression of existing NTLs, which may be due to an important progression in one lesion only or in several lesions
Complete response:	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10mm short axis).
Non-CR/Non-PD:	Persistence of one or more NTLs with no evidence of progression.
Not evaluable:	Only relevant when one or some of the NTLs have not been assessed and in the Investigator's opinion they are not able to provide an evaluable overall NTL assessment.
Not applicable:	Only relevant if there are no NTLs at baseline

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

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

If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present and should be treated as NE in the derivation of overall visit response.

3.2.3 Overall visit response

Table 5 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 5 Overall visit response

Target	Non-target	New lesions	Overall visit response
CR	CR (or NA)	No	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	Non-PD or NE	No (or NE)	PR
SD	Non-PD or NE	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NE	Non-PD	No	NE
NA	CR	No	CR
NA	Non-CR/Non-PD	No	SD
NA	NE	No (or NE)	NE
NA	Non-PD	NE	SD

3.2.4 Blinded Independent Central Review (BICR) Assessment

A planned BICR of a random sample of scans, from approximately 200 evaluable patients (with both progressive and non-progressive disease by investigator assessment), used in the assessment of tumours using RECIST 1.1 will be conducted. However, imaging assessments for all patients will be collected in case BICR of additional patients is required at a later date. All imaging assessments including unscheduled visit scans will be collected on an ongoing basis and sent to an AstraZeneca appointed CRO to enable central analysis.

The independent review charter contains the details of the independent central review conducted by AstraZeneca-appointed central . For each patient, the independent reviewer will provide time point response data and the relevant scan dates for each time point with supporting measurements and assessments. The time response data with the relevant scan dates from the independent review of scans will be combined programmatically to derive date of progression.

RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression will be determined based on the earliest of the scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of either reviewer where both select PD as time point response and there is no adjudication.

Results of this independent review will not be communicated to Investigators and the management of patients will be based solely upon the results of the RECIST 1.1 assessment conducted by the Investigator.

In order to ensure 200 evaluable patients have BICR data available, 220 patients will be identified to be reviewed by BICR. This has been chosen to allow a 10% overage for any selected patients who may not be evaluable by central review, for example because of poor imaging quality. All identified patients who have evaluable data will be included in summaries and listings of BICR data, even if this results in more than 200 evaluable patients.

To generate the random sample of 220 patients, a simple random sampling approach will be applied. The proportion of patients with PD compared to non-PD will not be controlled using this method of patient selection. BICR of sampled patients will be completed before data-base lock for the primary analysis of PFS. An audit plan is included in Appendix B to outline the criteria for determining when a complete review of scans will be conducted via a central review.

Further details are provided in section 4.2.6.

3.2.5 Best objective response (BOR)

BOR will be calculated based on the overall visit responses obtained up until RECIST 1.1 progression is documented. In the absence of RECIST 1.1progression, BOR is determined using visit responses up until the last evaluable overall visit response. This will be irrespective of whether patients discontinued treatment. Tumour assessments performed after the start of subsequent therapy will not be included in the calculation of BOR.

For patients who die in the absence of progression and have no evaluable overall visit responses prior to death, set the BOR to PD.

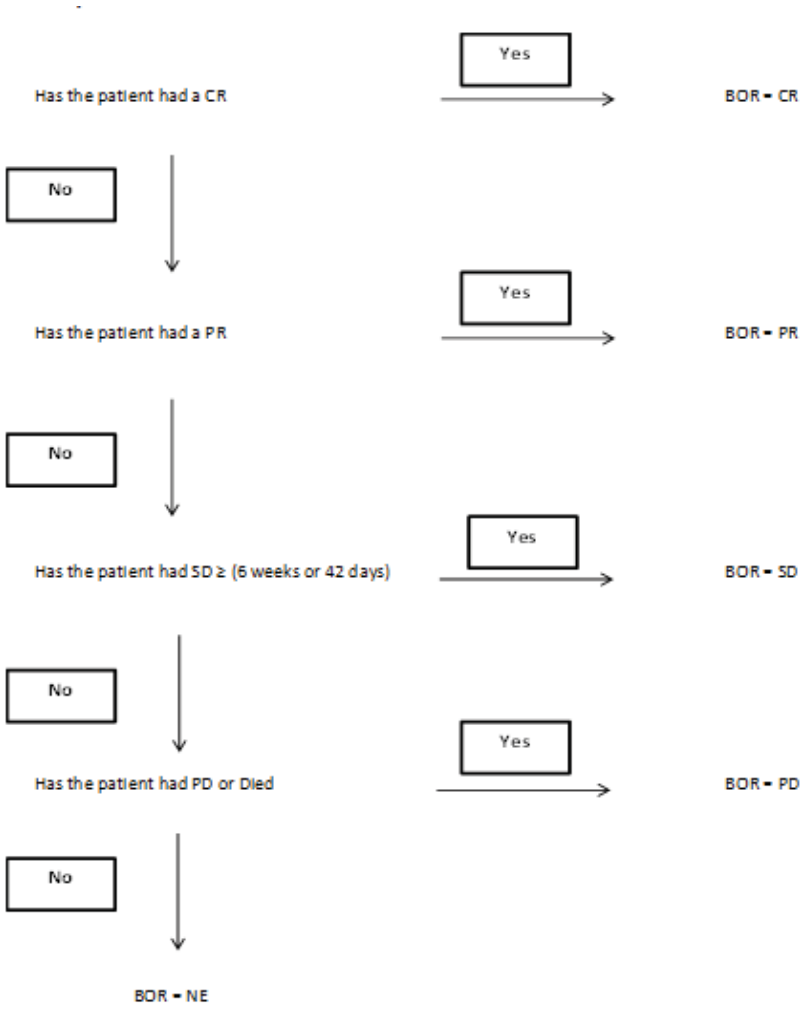
A patient's overall best objective response will be determined as follows:

- CR: At least one visit response of CR.
- PR: If not CR, at least one visit response of PR.
- SD≥ 6 weeks: If not CR or PR, stable disease recorded at least 42 days after randomisation (6 weeks from randomisation with no visit window allowed).
- PD: Progression, or death in the absence of CR/PR or SD.

- NE: No evidence of CR/PR or SD or PD or death.
- For patients who progress and subsequently have a response, then the best objective response is only derived from assessments up to and including the time of the progression (i.e., it will not include the response after the patient has progressed).
- Measurable disease according to RECIST 1.1 (i.e., the presence of TLs) is one of the inclusion criteria of the study. If, however, a patient does enter the study without measurable disease, then they will be assessed for objective tumour response based on their non-target lesions and any new lesions that they may have.

When programmatically deriving BOR the programs should derive the response in order of checking for best outcome to worst (Figure 2)

Figure 2 Flow diagram for determining best objective response



3.3 Primary efficacy outcome variable

3.3.1 Progression Free Survival (PFS)

The primary efficacy outcome variable of this study is PFS (defined by RECIST 1.1 as assessed by the investigator).

PFS is defined as the time from randomisation until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to progression.

Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after two or more missed visits, the patient will be censored at the time of the latest evaluable RECIST1.1 assessment. Two missed tumour assessment visits is defined as no evaluable tumour assessment within 14 weeks (98 days) of randomisation or the previous evaluable RECIST 1.1 measurement. If the patient has no evaluable visits or does not have baseline data they will be censored at day 1 unless they die within two visits of baseline.

PFS (days) = Event date – randomisation date + 1.

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

RECIST 1.1 assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- Date of progression will be determined based on the earliest of the dates of the component that triggered the progression
- When censoring a patient for PFS the patient will be censored at the latest of the dates contributing to a particular overall visit assessment

3.4 Secondary efficacy outcome variables

3.4.1 Overall survival (OS)

OS is defined as the time from the date of randomisation until death due to any cause.

OS (days) = Death date or Censor date – randomisation date + 1.

Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive.

For patients who have not died before the Data Cut-off (DCO), this study will endeavour to ensure that all patients are followed up for their survival status once the DCO is known. If there are no survival or death data available after the DCO, then the last date the patient is

known to be alive will be calculated from the last assessment date of all modules (except the visit module) on the database on or before the DCO.

Note, survival calls will be made in the two weeks following the date of DCO for the analysis, and if a patient is confirmed to be alive or has died up to 14 days after the DCO, then the patient will be censored for OS on the date of the DCO.

All efforts will be made to collect data on exact death date. In case there is an occasional need to impute a partial death date, the following rules will be applied. Date of death will be imputed only if month and year are available, following this convention:

- If no data are available in the month/year that indicate the patient was alive in that month, impute to 15th
- If patient has data in the month/year, impute to the day following the last available date which indicates patient was alive
- If only year is available, the date should not be imputed, patient will be censored at last known date to be alive in analysis

3.4.2 Objective response rate (ORR)

ORR is defined as the proportion of patients with at least one visit response of CR or PR.

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. This will be irrespective of whether patients discontinued treatment. For the patients who received a subsequent therapy, the tumour assessments performed after the start of subsequent therapy will not be included in the calculation of ORR. As measurable disease is an inclusion criterion for the study, the denominator for ORR will be all randomised patients.

A confirmed response is defined as a response (CR or PR) followed by a visit of response (CR or PR) at the next scheduled RECIST 1.1 assessment. Confirmation of response is not required for declaring a PR or CR in the overall ORR, but the number of confirmed response will also be tabulated.

Objective tumour response will be calculated from best objective response. Objective response will be derived as no/yes (0/1) variable. Patients with a BOR of CR or PR will be assigned ‘Yes’. Patients not having a BOR of CR or PR will be assigned ‘No’.

3.4.3 Duration of response (DOR)

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

If a patient does not progress following a response, then their duration of response will be censored at the PFS censoring time. Duration of response will not be defined for those patients who do not have documented response.

3.5 Safety and tolerability outcome variables

Safety data will not be formally analysed. Data from all cycles of randomised treatment will be combined in the presentation of safety data. 'On treatment' will be defined as assessments between date of first dose and 30 days following last dose date of selumetinib/placebo.

The Safety analysis set will be used for reporting of safety data.

3.5.1 Adverse event

The definitions of adverse events (AEs) and serious adverse events (SAEs) are given Sections 6.4.1 and 6.4.2 of the CSP.

The Medical Dictionary for Regulatory Activities (MedDRA Version 18.1) dictionary will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 4.03).

Any AE occurring before treatment or more than 30 days after treatment with selumetinib/placebo will be included in the data listings but will not be included in the summary tables of AEs.

Any AE occurring within 30 days of study treatment discontinuation (i.e., the last dose date of selumetinib/placebo) will be included in the AE summaries. Any events in this period that occur after a patient has received further therapy for cancer (following discontinuation of study medication) will be flagged in the data listings.

Other significant adverse events (OAE)

An AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and AEs leading to discontinuation, AEs of special interest, AEs with CTC grade 3 or higher, AEs causally related to study drug, AEs with an outcome of death or AEs leading to dose reduction, interruption, modification, dose delay. Based on the expert's judgement, significant adverse events of particular clinical importance, after consultation with the Global Patient Safety Physician, may be considered OAEs and reported as such in the CSR. A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

AEs of special interest

Some clinical concepts (including Grouped AEs and some selected individual preferred terms) are considered "AEs of special interest" (AESI) to the selumetinib program. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AESI's and identified higher level and preferred terms that contribute to each AESI (see appendix E). Further reviews will take place prior to Database Lock (DBL) to ensure any further terms not already included are captured within each AESI.

3.5.2 Laboratory

Blood and urine samples will be used for determination of clinical chemistry, haematology and urinalysis parameters. Clinical chemistry, haematology and urinalysis will be taken at all scheduled visits. The laboratory parameters to be collected are given in the protocol Section 6.4.5.5, Table 4 of the CSP.

Change from baseline in haematology and clinical chemistry variables will be calculated for each post-dose visit on treatment. CTC grades will be defined at each visit according to the CTC grade criteria using local or project ranges as required, after conversion of lab result to corresponding SI units. The following parameters have CTC grades defined for both high and low values: Potassium, Sodium, Magnesium and Corrected calcium so high and low CTC grades will be calculated. CTC grades are not defined for Total protein, Urea, Absolute eosinophil count, Absolute basophil count, Absolute monocyte count.

Corrected calcium: Corrected Calcium and Calcium Phosphate product will be derived during creation of the reporting database using the following formulas:

Corrected calcium = Total calcium (mmol/L) + ([40 – Albumin (G/L)] x 0.02)

Calcium Phosphate (mmol/L) = Corrected Calcium (mmol/L) x Phosphate (mmol/L)

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range) and high (above range).

The maximum or minimum on-treatment value (depending on the direction of an adverse effect) will be defined for each laboratory parameter as the maximum (or minimum) post-dose value at any time (see Appendix A).

Project reference ranges will be used for the primary interpretation of laboratory and vital signs data, with the exception of Alkaline Phosphatase (ALT), Aspartate Aminotransferase (AST) and total bilirubin for which the local laboratory ranges will be used for the primary interpretation. The project reference ranges will be based on the AstraZeneca reference ranges where they exist; otherwise project reference ranges agreed by the Global Safety Physician and the study physician will be used. The range used for each parameter will be listed (see Appendix C).

The denominator used in laboratory summaries of CTC grades will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality.

For example:

- If a CTCAE criterion involves a change from baseline, evaluable patients would have both a pre-dose and at least 1 post-dose value recorded.

- If a CTCAE criterion does not consider changes from baseline, to be evaluable the patient need only have 1 post dose-value recorded.

3.5.3 Vital signs

Vital sign assessments (resting blood pressure (BP) and pulse rate), including weight, will be performed as per the study plan and at the time of any echocardiogram assessment. Height will be assessed at screening only.

Change from baseline in vital signs variables will be calculated for each post-dose visit on treatment.

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range) and high (above range). All values classified as high or low will be flagged on the listings.

The denominator in vital signs data should include only those patients with recorded data.

3.5.4 Electrocardiograms (ECG)

Twelve-lead ECG will be performed as per the study plan. Patients should be supine and at rest 10 minutes prior to recording the ECG.

At each time point the Investigator's assessment of the ECG (normal, borderline or abnormal) and heart rate, duration of QRS complex, RR, PR and QT intervals will be collected. QTcF (Fridericia) will be calculated programmatically using the reported ECG values (RR and QT).

$$\text{QTcF (msec)} = \frac{\text{QT (msec)}}{\sqrt[3]{\text{RR (sec)}}}$$

For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point.

3.5.5 Echocardiogram

Left ventricular ejection fraction (LVEF), end diastolic and end systolic left ventricular volumes will be recorded at each echocardiogram assessment as per the study plan. If an echocardiogram scan cannot be taken, a MUGA scan to assess LVEF will be conducted.

If there are a reasonably large number of post-baseline observations, the change from baseline in LVEF will be calculated for each post-dose visit.

3.5.6 Physical examination

A physical examination will be performed as per the study plan and physical examination (general appearance, skin, head and neck, lymph nodes, thyroid, musculoskeletal/extremities, cardiovascular, respiratory, abdomen, and neurological) will be evaluated as normal/abnormal at assessment.

3.5.7 Ophthalmic assessments

An ophthalmologic examination (best corrected visual acuity, intraocular pressure and slit lamp fundoscopy) should be performed by an ophthalmologist at baseline and on occurrence of an AE of visual disturbance and at the 30 day follow-up visit, if patient had abnormal examination at discontinuation visit. Best corrected visual acuity for near vision and distance vision may be recorded using different notation due to local site conventions. The eCRF module only provides room to record one notation type. If the Snellen notation is used for one of the assessments then the notation will only be either Snellen fraction foot or Snellen fraction metre and the notation for the other assessment will not be displayed.

3.5.8 Duration of exposure

Selumetinib/placebo

Total exposure to selumetinib/placebo

Total exposure to selumetinib/placebo will be the 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 selumetinib/placebo

Actual exposure to selumetinib/placebo will be the time (days) from the 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 on dose interruption i.e. number of days with dose = 0 mg

Missed doses will not be captured in eCRF (electronic Case Report Form). The actual exposure calculation makes no adjustment for dose reductions that may have occurred.

Docetaxel

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

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

Patients who permanently discontinue during a dose interruption

If a patient permanently discontinues study treatment during a dose interruption, then the date of last administration of study medication recorded on DOSDISC will be used in the programming. The dose interruption will not be included as a dose interruption in the summary tables but will be recorded on the DOSE module and consequently appear in the listing for dosing.

3.5.9 Dose intensity

Relative dose intensity (RDI) and Percentage intended dose (PID) for selumetinib/placebo and docetaxel will be defined as follows:

- RDI = 100% * 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 defined as 0 days for selumetinib/placebo and 20 days for docetaxel.

- PID = 100% * 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.

RDI and PID for selumetinib/placebo and docetaxel will be calculated for the entire intended treatment period (censored at data cut-off). RDI and PID for docetaxel will also be calculated over defined number of cycles (censored at the earliest of data cut-off or end of the planned number of cycles). The planned number of cycles used will be 6.

For both RDI and PID progression and censoring events will be based on the primary analysis definition of PFS.

3.6 Pharmacokinetic variables

The final PK analyses will be the responsibility of
AstraZeneca, UK.

The plasma concentration-time data for selumetinib and N-Desmethyl selumetinib will be analysed by mixed effects modelling. The aim is to evaluate the pharmacokinetic characteristics of selumetinib and N-Desmethyl selumetinib including estimating pharmacokinetic variables and quantifying variability. An attempt may be made to identify potentially important covariates such as weight, age, sex and/or concomitant medications. The relationships between plasma exposure and pharmacodynamic (PD) parameters (efficacy and safety) will be explored. Initially graphical representation of the data will be performed but mixed effects modelling of the data may be carried out if feasible.

Other metabolites of selumetinib (e.g. selumetinib amide) may also be analysed as described above. A population PK analysis will be carried out. A detailed pharmacokinetic analysis plan will be provided prior to database lock (DBL).

The analysis will be reported in a separate pharmacokinetic report outside of the CSR.

3.7 Symptom endpoints and Health-related quality of life (HRQoL)

Symptoms will be assessed using symptom improvement rate and time to symptom progression (TSP) based on the Average Symptom Burden Index (ASBI) (loss of appetite, fatigue, cough, dyspnoea, hemoptysis, and pain) from the LCSS.

HRQoL will be assessed using the summary items of the LCSS (symptom distress, interference with activity levels and global HRQoL), the average LCSS score (the mean of all 9 items of the LCSS), and the scores for each of the 8 health domain scales and the 2 summary measures of the 36 Item Short-Form Health Survey Version 2 (SF-36v2).

If a patient is unable to complete the questionnaires unaided, the designated clinic staff member will read the questions verbatim to the patient and record the answers, without interpretation. Where significant assistance in completing the questionnaires is required, this must be recorded. Patient's HRQoL score will not differ if clinic staff records the answers.

If there are cases in which more than one questionnaire has been completed on the same day, and provide different answers, the questionnaire with the worst case approach will be used: the highest score will be used for the LCSS and ASBI and the lowest for the SF-36v2.

3.7.1 Lung Cancer Symptom Scale (LCSS)

LCSS evaluates six major symptoms associated with lung malignancies and their effect on overall symptomatic distress, functional activities, and global quality of life. It captures the symptoms most likely to be influenced by therapeutic interventions. The LCSS includes nine visual analogue scales (VAS) and has a recall period of 'the past day', which has been operationalized as the past 24 hours. Six of the nine items address major symptoms of lung cancer and constitute the ASBI; loss of appetite, fatigue, cough, dyspnoea, haemoptysis, and pain, while the remaining three VAS are summary items which assess symptom distress, interference with activity levels, and global quality of life.

The interval level VAS version uses 100 mm lines to measure the intensity for each item. Each item is given a score equal to the length of the line marked by the patient with scores ranging from 0- 100, and with 0 = the best possible status, and 100 = to the worst possible status. Higher score indicating greater symptom burden. Seven of the items have the anchors 'none'/'not at all' and 'as much as it could be'/'as bad as it could be'. The appetite item ranges from 'as good as it could be' to 'as bad as it could be' and the global HRQoL item ranges from 'very high' to 'very low'.

Patient-reported symptoms will be assessed using the ASBI, which is sub-score of the LCSS. The ASBI score is derived from the mean of the scores from the above mentioned six individual symptom questions of the LCSS in accordance with the recommendations of the developers.

Baseline

Baseline will be defined as the last non-missing LCSS assessment prior to randomisation for symptoms and summaries.

Missing data

Missing items will not be imputed. If at least one item is missing then the overall mean LCSS score (mean of all 9 items) for that patient for that visit is missing. However, if items 1 to 6 (loss of appetite, fatigue, cough, dyspnoea, hemoptysis, and pain) are not missing, then the ASBI can be calculated.

Categories of responses of LCSS

The responses to the LCSS at each assessment will also be categorised as improved, and stable based on the changes from baseline. At a given visit, the criteria for a relevant change also described as the minimum clinically important difference (MCID) (Hollen and Gralla 2000; Royston and Parmar 2011
Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Stat Med 2011 30:2409-2421

Sarna et al 2008 and De Marinis et al 2008) will be used to assign a visit response for each score (see Table 6**)**

Table 6 MCID for average, ASBI and individual LCSS score: Visit score

	Visit response [a]		
Score	**Improved**	**Worsened**	**Stable**
Average score [b]	≤ - 10	≥ + 10	Otherwise
ASBI score [c]	≤ - 10	≥ + 10	Otherwise
Individual score	≤ - 15	≥ + 15	Otherwise

[a] Based upon the visit response derived from change from baseline.
[b] Mean of all 9 individual symptom questions.
[c] Mean of the scores from the 6 individual symptom questions (loss of appetite, fatigue, cough, dyspnoea, haemoptysis and pain)

Some patients may have baseline scores that are too close to the minimum or maximum score to allow an improvement or a deterioration to occur. For example, LCSS has minimum score 0 (good) and maximum score 100 (bad). For ASBI, patients with a baseline score of < 10 would not be able to show an improvement (decrease in ASBI of ≥ 10); similarly patients with a baseline score > 90 would not be able to show deterioration (increase in ASBI ≥ 10). These patients will be included in visit level sum aries but will be excluded from the denominator for symptom improvement rate or time to symptom progression analyses, as appropriate.

For ASBI, these patients should be included in the visit level summaries of improved/worsened /stable as follows:

- Patients with a baseline score < 10:
  - A visit response of 'improvement' is not possible for these patients. Visit responses of worsened and stable will be calculated as usual. If they improve to a score of 0 this will be classed as 'stable'.
- Patients with a baseline score of > 90:
  - A visit response of 'worsened' is not possible for these patients. Visit response of improved and stable will be calculated as usual, however if a patient deteriorates to a score of 100 their visit response would be NE (as we cannot rule out that a worsening could have happened. If the score is ≤ 99 the visit response would be 'stable'.

For other LCSS variables, the values used for the relevant changes and baseline scores will be as described in Table 6. Example, for individual scores, where improvement or worsening is a change of ≥15, patients with a baseline value < 15 would not be able to show an improvement; similarly patients with a baseline score > 85 would not be able to show a deterioration. The rules applied to ASBI above would be adapted and applied to the individual scores (using 85 instead of 90 and 15 instead of 10).

3.7.1.1 Time to symptom progression (TSP)

Symptom changes will be determined based on changes in the ASBI score compared to baseline, with a minimum clinically meaningful change in symptoms defined as a change the ASBI score of ≥10 (Hollen and Gralla 2000; Royston and Parmar 2011
Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Stat Med 2011 30:2409-2421

Sarna et al 2008).

The ASBI is considered to target the most relevant and important symptoms and time to symptom progression (TSP) will be assessed through evaluation of the change in ASBI between baseline and later time points.

TSP will be defined as the time from randomisation until the date of first clinically meaningful symptom deterioration (an increase in the ASBI from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, provided death occurs within two LCSS assessment visits of the last LCSS assessment where ASBI could be evaluated, and regardless of whether the patient withdraws from randomised therapy or receives another anticancer therapy prior to symptom deterioration.

Patients whose symptoms (as measured by ASBI) have not shown a clinically meaningful deterioration (defined as a decrease in the ASBI from baseline as ≥10) and who are alive at the time of the analysis will be censored at the time of their last LCSS assessment where ASBI could be evaluated. Also, if symptoms progress after two or more missed LCSS assessment visits or the patient dies after two or more missed LCSS assessment visits, the patient will be censored at the time of the last LCSS assessment where ASBI could be evaluated.

Two missed visits will be defined as no assessments within 8 weeks (56 days) of randomisation or the previous evaluable assessment. This timeframe will be used even during cycle 1 when LCSS is assessed weekly.

If a patient has no evaluable visits or does not have baseline data they will be censored at day 1.

TSP (days) = Event date – randomisation date + 1.

The population for analysis of TSP will include a subset of the FAS population who have baseline ASBI scores ≤ 90. For the individual symptoms, the population will include a subset of the FAS population who have a baseline individual symptom score ≤ 85 for the relevant symptom.

In the primary analysis, RECIST 1.1 progression will not be considered as progression of symptoms and data will not be affected by RECIST progression. However two sensitivity analyses will be performed:

1. Where RECIST 1.1 progression is considered as an event of symptom progression.

If a patient has both RECIST 1.1 progression and symptom progression, the earliest date of progression will be used. RECIST 1.1 progression must occur before or within 8 weeks (56 days) of the last evaluable ASBI assessment to be considered as symptom progression.

2. Where patients are censored at RECIST 1.1 progression if symptom progression has not yet occurred.

If a patient has symptom progression before RECIST 1.1 progression then the earlier symptom progression date will be used. If a patient is censored for symptom progression before RECIST 1.1 progression occurs, then the earlier date of censoring will be used. If a patient has RECIST 1.1 PD before either symptom progression or censoring for symptom progression then the patient will be censored at the earlier RECIST 1.1 progression date.

3.7.1.2 Symptom Improvement rate

A clinically meaningful improvement in symptoms will be defined as a decrease in the ASBI from baseline of ≥10.

The symptom improvement rate will be defined as the number (%) of patients with two or more consecutive assessments at least 18 days apart (i.e. 21 days allowing a visit window of 3

days) which showed a clinically meaningful improvement in symptoms from baseline. The denominator consisting of a subset of the FAS population who have baseline ASBI scores ≥10.

This analysis will be repeated for the individual symptoms, in this case the denominator will include a subset of the FAS population who have a baseline individual symptom score ≥ 15 for the relevant symptom.

3.7.2 Short-Form Health Survey (SF-36v2)

Patient-reported HRQoL will be assessed using the SF-36v2 questionnaire.

The SF-36v2 has a 1-week recall period. The SF-36v2 will be completed after the LCSS but before any study related assessments at the time points (i.e. baseline, cycle 3, cycle 5, objective disease progression and 30 days post-progression). If patient discontinues study treatment for reasons other than objective disease progression, SF-36v2 is to be completed at the time of objective disease progression (regardless of whether the patient has received another anti-cancer therapy prior to progression) and again approximately 30 days later.

It assesses HRQoL/health status using multi-item scales to measure the following eight dimensions: physical functioning (10 items), role limitations due to physical health problems (4 items), bodily pain (2 items), social functioning (2 items), general mental health (5 items), role limitations due to emotional problems (3 items), vitality, energy or fatigue (4 items) and general health perceptions (5 items). In addition, psychometrically-based physical component summary (PCS) and mental component summary (MCS) scores can also be calculated. Both the summary scores and the domain scores have well established evidence of validity and reliability across diverse patient groups (Ware et al 1993; McHorney et al, 1993; McHorney et al, 1994).

The scores will be derived in accordance with the SF-36v2 manual and interpretation guide. The SF-36v2 scale scores are scored so that a higher score indicates better health and are based on the sum of the items included in a given scale, transformed to a 0-100 scale.

3.7.2.1 Health domain scales and Physical and mental component summary scores

The items included in each scale are as follows:

- PF: Physical functioning - Question 3 (3a-3j)
- RP: Role limitations due to physical health problems – Question 4 (4a-4d)
- BP: Bodily pain – Questions 7 and 8
- SF: Social functioning – Questions 6 and 10
- MH: General mental health – Question 9 (9b, 9c, 9d, 9f, 9h)

- RE: Role limitations due to emotional problems – Question 5 (5a-5c)
- VT: Vitality, energy or fatigue – Question 9 (9a, 9e, 9g, 9i)
- GH: General health – Question 1 and 11 (11a-11d)

The two summary scores PCS and MCS will also be calculated in accordance with the SF-36v2 manual.

The absolute values and change from baseline will be calculated for each domain scales at each scheduled post-baseline assessment.

3.7.3 PRO compliance rates

Summary measures of overall compliance and compliance over time will be derived for LCSS and SF-36v2. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.
- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not 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 no missing items (i.e. all items for the overall LCSS, items 1 to 6 for the ASBI score).
- Overall PRO compliance rate is defined for each randomised treatment group as: Total number of evaluable questionnaires across all time points, divided by total number of questionnaires expected to be received across all time points multiplied by 100.
- Overall patient compliance rate is defined for each randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the time point (as defined above), divided by number of patients still expected to complete questionnaires. Similarly the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires. Compliance rate is summarised using scheduled visits. Visit windowing only applies to summaries of PROs over time

3.8 Exploratory variables

3.8.1.1 Health state utility

Preference based health status utility values are often used in economic evaluations to estimate the health benefit of treatments. These utility values will be derived from the SF-36v2 non-preference based generic PRO instrument. Two approaches will be used to generate utility values:

- Reduce the SF-36v2 to the SF-6D and apply an algorithm to derive utility value (Brazier et al.2002)
- Mapping from the SF-36v2 to the EQ-5D using a model specification as outlined in Rowen et al 2009

A detailed analysis will be documented in the payer analysis plan. The respective analysis will be reported separately from the CSR.

3.8.2 Healthcare Resource Use

Frequency and estimates of healthcare resource use, including hospital episodes, type of contact (hospitalisation, outpatient, day case), reason, length of stay by ward type (including ICU), concomitant medication and procedures and tests undertaken will be derived from the resource use information. The tables will be produced using the FAS.

A payer analysis plan will described the detailed analysis. The analysis will be reported separately from the CSR.

4. ANALYSIS METHODS

4.1 General principles

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

- Descriptive statistics will include number of non-missing patients (n), arithmetic mean, standard deviation, median, minimum, and maximum values for continuous variable, and frequencies and percentages for categorical variables.
- For continuous data, mean, standard deviation and median will be rounded to 1 additional decimal place compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.
- Categorical variables will be summarised by frequency counts and percentages for each category.
- For categorical data, percentages will be rounded to 1 decimal place.

- Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment group according to the particular analysis set.
- Stratification data from Interactive Voice Response System (IVRS) randomization will be used.
- SAS® version 9.1.3 or higher will be used for all analyses.

There will be one treatment comparison of interest:

- Selumetinib 75mg bd in combination with docetaxel 75mg/m^2 vs placebo in combination with docetaxel 75mg/m^2

The formal statistical analysis will be performed to test the following hypotheses:

- H_0: There is no difference between Selumetinib 75 mg + docetaxel 75 mg/m^2 and placebo + docetaxel 75 mg/m^2
- H_1: There is a difference between Selumetinib 75 mg + docetaxel 75 mg/m^2 and placebo + docetaxel 75 mg/m^2

Results of all statistical analysis will be presented using a 95% CI and 2-sided p-value, unless otherwise stated.

4.2 Analysis methods

Table 7 gives all formal statistical analyses planned for this study:

Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoint Analysed	Analyses
PFS	Primary analysis: Stratified log-rank test based on RECIST 1.1 data from investigator tumour assessment Sensitivity analyses: 1) Evaluation Time bias 2) Attrition bias 3) Ascertainment bias Secondary analyses: Cox proportional hazards model[b] Subgroup analyses[a]: Cox proportional hazards model

Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoint Analysed	Analyses
Overall survival	Primary analysis: Stratified log-rank test Secondary analyses: Cox proportional hazards modelb Subgroup analyses (same as PFS)
Objective response rate	Logistic regression using investigator's assessment of RECIST 1.1
Time to symptom progression (ASBI)	Stratified log-rank test Sensitivity analyses: Attrition bias Analysis where RECIST 1.1 PD is an event Analysis where censoring at RECIST 1.1 PD
Symptom improvement rate (ASBI)	Logistic regression
Time to symptom progression (Six individual symptoms from ASBI)[c]	Stratified log-rank test
Symptom improvement rate (Six individual symptoms from ASBI) [c]	Logistic regression

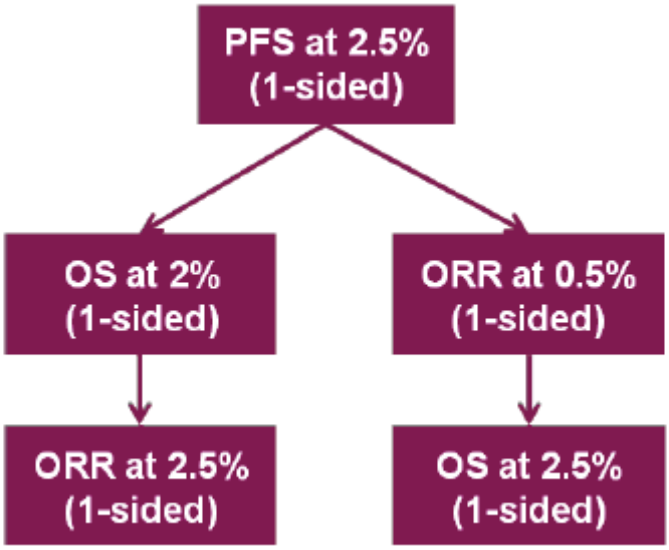
[a] Only hazard ratio and confidence intervals (CI) will be presented for subgroup analyses (no p-values).
[b] An initial model will be constructed, containing treatment and the two stratification factors alone to assess the effect of covariates on the HR estimate.
[c] Only odds ratio/hazard ratio and confidence intervals will be presented (no p-values).

4.2.1 Multiple testing strategy

In order to describe the nature of the benefits of selumetinib treatment, PFS, OS, ORR, time to symptom progression and symptom improvement rate will be tested at a two-sided significance level of 5%.

However, in order to strongly control the type I error at 2.5% 1-sided, a multiple testing procedure (MTP) with an alpha-exhaustive recycling strategy (Burman et al 2009) will also be employed across the PFS and secondary endpoints OS and ORR. With this approach, the primary endpoint PFS and secondary endpoints OS and ORR will be tested in a pre-defined order as shown in Figure 3 below:

Figure 3 Multiple testing procedures for controlling type-1 error rate



The primary endpoint (PFS) is tested at a 1-sided α of 2.5%. If the primary hypothesis of PFS is rejected for superiority, the secondary endpoints will then be tested in the MTP using a weighted proportion of alpha (test mass; the total test mass equals alpha) and test mass becomes available after each rejected hypothesis, which is recycled to secondary endpoints not yet rejected. This testing procedure stops when the entire test mass is allocated to nonrejected endpoints. Implementation of this pre-defined ordered testing procedure, including recycling, will strongly control the Type I error at 2.5% (1-sided) amongst the primary (PFS) and key secondary endpoints.

4.2.2 Progression Free Survival

The primary analysis will be based on the programmatically derived PFS based on investigator-recorded assessments and will compare the PFS between selumetinib in combination with docetaxel versus placebo in combination with docetaxel.

The primary analysis of PFS will be based on the FAS. PFS will be analysed using a stratified log-rank test using the Breslow method to handle ties (Breslow, 1974). The stratification factors are WHO Performance Status (0/1) and Histology (squamous/non-squamous). The following variables will be used in stratified log-rank test:

- PFS: time to event (days)
- Censoring indicator variable:1=censored; 0=event
- Treatment:1= Selumetinib + Docetaxel; 0= Placebo + Docetaxel
- WHO Performance Status: 0= 0 and 1= 1

- Histology: 1= squamous; 0= non-squamous.

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

To identify squamous/non- squamous status following codes were used in the eCRF:

Histology type 401: squamous

Histology type 0 or 99 or 407 or 412 or 415: non-squamous

The effect of treatment will be estimated by the HR together with its corresponding 95% and p-value. The HR and its CI can be estimated from the stratified log-rank as follows (Anderson et al 2004
Anderson P.K, Hansen, M.G., Klein, J.P. Regression analysis of restricted mean survival time based on Pseudo-observations. Lifetime Data Anal. 2004 Dec:10(4): 335-350

Berry et al.1991, Collett, 2003,Sellke and Siegmund 1983):

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

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

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

Kaplan-Meier (KM) plot of PFS will be presented by treatment group. Summaries of the number and percentage of patients experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided along with median of PFS for each treatment.

The proportion of patients alive and progression free at 6 months (183 days) and 12 months (365 days) will be summarised (using the KM curve) and presented by treatment group.

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

The final analysis of PFS will take place on a pre-specified date when it is predicted that 332 death events will have occurred. The exact date will be predicted by modelling the blinded death rate data.

When conducting the stratified analyses, if a WHO Performance Status by Histology status stratum for any treatment arm contains less than 5 events, the analyses will be stratified by WHO Performance Status only.

The final strategy selected will be applied to all other stratified time to event endpoints and analyses.

For stratified treatment analyses, the IVRS value will be used to be consistent with the randomization scheme.

Additional supportive summaries/graphs

In addition, the number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest scan prior to the data cut-off was more than one visit (7 weeks) prior to the data cut-off date.

A summary of the number of patients who have incomplete PFS follow-up, i.e. were alive and progression free at the date of data cut-off and had not had a RECIST 1.1 assessment within 7 weeks of the data cut-off will be provided by treatment group. Additionally, summary statistics will be given for the number of days from censoring to data cut-off for all censored patients.

A summary of the median duration of follow-up will also be presented. Each patient's duration of follow-up will be defined as the number of days from randomisation to the date ofcensoring (date last known to be non-progressor) in censored (not progressed) patients only.

All of the collected RECIST 1.1 data will be listed for all randomised patients. In addition, a summary of new lesions (i.e., sites of new lesions) will be produced.

Proportionality assumption

The assumption of proportionality will be assessed. Proportional hazards will be reviewed by examining plots of complementary log-log (event times) versus log (time) and by fitting an extended Cox model that includes Time by Treatment interaction. If a lack of proportionality is indicated (for example due to the presence of a statistically significant treatment by time interaction or the lines on the log-log plots are not reasonably parallel), the HR from the primary analysis can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves.

The treatment effect will also be described by presenting piecewise HRs calculated over distinct time-periods of 0-3, 3-6, >6 months (0-91, 92-183, 184+ days). The piecewise model will be implemented by the addition of a time varying covariate (based on the periods in the

previous sentence) as per Collett, 2003. In short the piecewise Cox model divides the time period in the primary analyses into three intervals, namely (0, t1), (t1, t2) and (t2, inf). Letting X be an indicator variable associated with the two treatments, where X = 0 for placebo and X = 1 for selumetinib. The piecewise Cox regression model is then fitted by defining three time dependent variables, $Z_1(t)$, $Z_2(t)$ and $Z_3(t)$ as follows:

$Z_1(t) = 1$ if time in interval t0 to t1 and X = 1; 0 otherwise

$Z_2(t) = 1$ if time in interval t1 to t2 and X = 1; 0 otherwise

$Z_3(t) = 1$ if time > t2 and X = 1; 0 otherwise

The model for the hazard function for the ith individual at time t can be written as:

$$hi(t) = \exp (\beta_{1z1i}(t) + \beta_{2z2i}(t) + \beta_{3z3i}(t) + \beta_{4z}\text{WHOPS} + \beta_5\text{HISTOLOGY})h_0(t)$$

For each time period, the log(HR) for new treatment to standard treatment is:

Time period 1 = β_1

Time period 2 = β_2

Time period 3 = β_3

In addition, the restricted mean survival time (RMST) will be calculated and considered as a supportive analysis to aid the investigation and interpretation of the treatment effect (Royston and Parmar 2011). The RMST will be calculated using the pseudo-value approach (Anderson et al, 2004) and will be based on the maximum common survival time across the two treatment arms. Adjusted (based on the same covariates as included in the primary analysis model) RMST will be calculated using generalized estimating equations and the RMST reported for each treatment arm as well as the 95% CI and p-value.

4.2.2.1 Sensitivity Analyses for Primary Endpoint (PFS)

Sensitivity analyses will be performed to assess the possible presence of time-assessment bias (i.e. differential assessment times between treatment groups).

(a) Evaluation-Time bias

Sensitivity analyses will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled timepoints. The midpoint between the time of progression and the previous evaluable RECIST 1.1 assessment will be analysed using a stratified log-rank test. If the midpoint is found to be 0.5 of a day, this will be rounded down to the nearest whole day. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). Note that if the event contributing to the analysis is death in the absence of progression, then the time to event

remains unchanged (i.e. the time to death is not replaced by the midpoint between death and the previous evaluable RECIST assessment).

(b) Attrition bias

Attrition bias will be assessed by repeating the primary PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following two, or more, non-evaluable tumour assessments will be included. Two missed/ non-evaluable tumour assessment visits is defined as no assessment within 14 weeks (98 days) of randomisation or the previous evaluable RECIST 1.1 measurement. In addition, patients who take subsequent therapy (excluding radiotherapy) prior to progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring where the censoring indicator of the primary PFS analysis is reversed.

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

(c) Ascertainment bias

Ascertainment bias will be assessed from carrying out a blinded independent central review (see section 3.2.4 and 4.2.6). The objective of the BICR is to detect potential evaluation bias in the investigator assessment of PFS. Ascertainment bias will be assessed through the use of two different approaches:

- The first approach is a method proposed by Dodd et al 2011.
- The second approach is a parameter-free method proposed by Stone et al. 2015.

Therefore one table will be produced using the Dodd approach to present the extrapolated hazard ratio for the BICR in the full study and another table will be produced using the Stone approach to present the hazard ratio ratio between the BICR and the local evaluation (LE). For the Dodd approach if the upper limit of the one sided 97.5% CI is <1, then the BICR audit has confirmed the PFS effect per LE. For the Stone approach an upper limit of the one sided 90% CI around the ratio of hazard ratio per BICR and hazard ratio per LE being < 1.25 supports consistency of the local and BICR assessments. In both cases the analyses will be performed included the treatment effect only. More details of these analyses are given in Appendix B.

Another table will be produced presenting a concordance analysis between declared RECIST progression from the BICR and the LE. The categories that will be presented are:

- RECIST progression declared by both BICR and LE
- Progression date agreement (within 2 weeks)
- Progression date >=2 weeks earlier by BICR than by LE

- Progression date >=2 weeks earlier by LE than by BICR
- RECIST progression declared by LE but not by BICR
- RECIST progression declared by BICR but not by LE
- RECIST progression not declared by either BICR or LE

4.2.2.2 Subgroup analyses

The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors but from the results observed in Phase II (D1532C00016) it is not expected that these factors will be predictive factors for a qualitatively different treatment effect.

Subgroup analyses will be conducted comparing PFS between treatments in the subgroups of the FAS defined by the stratification factors WHO Performance Status (0/1) and Histology (squamous/non-squamous), plus the following factors:

- Gender: 1= Male; 0= Female
- Age at randomisation: 1= < 65 years; 0= > = 65 years
  - Patients with a missing age value will be included using the mean age (overall FAS) and categorised accordingly.
- Smoking status: 1= Smoker; 0= Non-smoker (never smoker)
  - Patients with a missing smoking status will be included in the 'smoker' category.
- Metastatic or Locally advanced disease: 1= Locally advanced; 0 = Metastatic
  - Patients with a missing disease status at screening will be included in the most common category of disease status. Patients with both locally advanced and metastatic disease at screening will be included in the 'metastatic' group.
- Race: 1= White; 0= Non-white
- Specific *KRAS* mutation subtype: 1= Codon 12/13; 0= Codon 61
- PD-L1 biomarker status: 1= PD-L1≥5%; 0= PD-L1<5%; 2= Unknown
  - Patients with a missing PD-L1 biomarker status will be in the Unknown category

  - If less than 25% of the patient samples fall in the greater than or equal to 5% category, then the cut-off will be changed from 5% to 1%

  - For patients with more than one PD-L1 result available the sample with the earliest assay performed date will be used to define PD-L1 status

- Region: 1= North America (Canada and US); 0= Other

- First line platinum therapy: 1= Yes; 0= No

For each subgroup, the HR and 95% CI will be calculated from a single Cox proportional hazards model that contains a term for treatment, the subgroup covariate of interest and the treatment by subgroup interaction term. The treatment effect HR will be obtained for each level of the subgroup from this model. The stratification factors will not be included in the model unless the stratification factor is the subgroup being analysed. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties. CIs will be profile likelihood CIs.

These hazard ratios and associated two-sided 95% CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

No adjustment to the significance level for testing will be made since all these analyses will be considered supportive of the primary analysis of PFS.

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 10 events in either treatment arm of the subgroup), the relationship between that subgroup and PFS will not be formally analysed. In this case, only descriptive summaries will be provided.

Other baseline variables may also be included if there is clinical justification or an imbalance is observed across the treatment arms.

4.2.2.3 Secondary analysis

Cox proportional hazards modelling will be employed to assess the effect of covariates on the HR estimate. The Cox model will be fitted using SAS® PROC PHREG with the Efron method to control for ties. CIs will be profile likelihood CIs. Before embarking on more detailed modelling, an initial model will be constructed, containing treatment and the two stratification factors alone, to ensure any output from the Cox modelling is likely to be consistent with the results of the stratified log-rank test.

Result from initial model and the model containing additional covariates will be presented.

Additional covariates for this model will be gender, age at randomisation, smoking, Metastatic or locally advanced disease, race, specific *KRAS* mutation subtype, PD-L1 biomarker status, region and 1st line platinum therapy.

The model will include the effect regardless of whether the inclusion of effect significantly improves the fit of the model providing there is enough data to make them meaningful.

4.2.2.4 Consistency of treatment effect between subgroups

The presence of quantitative interactions will be assessed formally 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 (stratification factors and subgroups mentioned in section 4.2.2.2), 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 a covariate does not have 10 events or more per treatment group per level (of the covariate) it will be included as a covariate in the model but the covariate-by-treatment interaction term will be omitted. If the fit of the model is not significantly improved then it will be concluded that overall the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% 2-sided 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.

4.2.3 Overall Survival (OS)

OS data will be analysed at the final analysis of PFS and will use the same methodology and model as described in section 4.2.2. Secondary analysis using cox proportional and subgroups will be performed. However, global interaction test will not be carried out for OS.

OS will be summarised for each arm using median and OS will be displayed graphically using KM plot as detailed for PFS.

A summary of survival status at the time of analysis will be produced. This will summarise the number of patients who have died, still in survival follow-up, are lost to follow-up (LTFU) and have withdrawn consent.

The final analysis of OS will be analysed when 332 deaths have occurred (~65% maturity). No further analyses of OS are planned beyond this point unless requested by Health Authorities.

Sensitivity analysis: Attrition bias

A Kaplan-Meier plot of the time to censoring where the censoring indicator of the primary OS analysis is reversed will be produced to assess whether there is an imbalance in censoring for OS.

Additionally, the number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if their survival status was not defined at the DCO.

Survival data will be listed for all randomised patients.

4.2.4 Objective response rate (ORR)

The ORR will be based on the investigator's assessment of RECIST 1.1. The ORR will be summarised for all patients in the FAS. A summary of ORR will be presented by treatment group. ORR will be compared between selumetinib in combination with docetaxel vs. placebo in combination with docetaxel using an adjusted logistic regression model, provided there are enough responses for a meaningful analysis. The model will include the stratification factors WHO Performance Status (0/1) and Histology (squamous/nonsquamous).

The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour selumetinib in combination with docetaxel) together with its associated 95% CI and 2-sided p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). CIs will be profile likelihood CIs (e.g. using the option 'LRCI' in SAS procedure GENMOD).

If there are not enough responses for a meaningful analysis (less than 10 responders in either treatment arm) using logistic regression then a Fisher's exact test using mid p-values will be presented.

The mid-p-value modification of the Fisher exact test amounts to subtracting half of the probability of the observed table from Fisher's p-value.

$$\text{Fisher's exact test mid p-value} = \text{Two sided p-value} - \frac{\text{Table probability}}{2}$$

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

4.2.4.1 Change in tumour size

The absolute values at baseline and week 6 and percentage change in target lesion tumour size from baseline at week 6 will be summarised using descriptive statistics. The number and percentage of patients in each treatment group whose week 6 data is imputed will also be presented.

The effect of selumetinib on percentage change in tumour size will be estimated from an analysis of covariance (ANCOVA) model including the percentage change in week 6 value as a response variable, a covariate for the treatment, the baseline tumour size and a covariate for the time from the baseline scan to randomisation. The model will also include the stratification factors. The number of patients, unadjusted mean and adjusted LSMEANS for each treatment group will be presented, together with the difference in adjusted LSMEANS, 95% CI and corresponding p-value.

If it is judged the data do not adequately follow a normal distribution based on the blinded data, then the use of log-transformed data or a non-parametric approach could replace the untransformed analysis as the primary approach.

Waterfall plots of best percentage change in tumour size for each treatment group will be produced.

Percentage change from baseline in tumour size at 6 weeks is based on RECIST target lesion measurements taken at baseline and at week 6. Tumour size is the sum of the longest diameters of the target lesions. Target lesions are measurable tumour lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to starting treatment. The percentage change in target lesion tumour size at week 6 will be obtained for each patient taking the difference between the sum of the target lesions at week 6 and the sum of the target lesions at baseline divided by the sum of the target lesions at baseline times 100. For log transformed data, the baseline scaled ratio in change in tumour size will be derived as Loge(Week 6 value/Baseline value).

Patients who progress before week 6 should have had a tumour assessment performed at the time of progression prior to treatment discontinuation. The tumour size from their latest progression assessment will be used instead of the week 6 assessment for these patients. If after imputation of missing week 6 lesion data as described below there remains a reasonable amount of missing tumour size measurement data, a non-parametric method will be used, assigning patients who have died with the worst rank.

Target lesion imputation:

For patients who have less than or equal to one-third of target lesions missing at week 6, assessment data from missing lesions may be scaled up proportionally to the sum of the corresponding lesions at baseline to give an estimated sum of diameters as described in section 3.2.

Apply a window around the week 6 visit:

Whenever tumour size data for the week 6 visit (Note: or visit at which progression was documented if before week 6) is available then this should be used in the analysis. A windowing rule will be applied and will follow the protocol allowed visit window; therefore any RECIST scan performed within ± 2 week of the protocol scheduled visit will be used for that visit.

If, after applying the above considerations to the missing data, there is still missing tumour size measurement data at week 6, the recommendation would be to follow the imputation process outlined below for each individual patient where data is missing (applied prior to blind review).

(a) If there are no tumour size data at week 6, but there are tumour size data collected at a visit prior to week 6 or the first visit after week 6, use all of the available data up to and including the first visit after week 6 (i.e., baseline and all visits up to and including the first visit after week 6) to fit a linear regression to the individual patient's baseline and follow-up assessment(s) to generate an estimated value for tumour size at week 6 and hence impute a change from baseline at week 6.

(b) If there is evidence of progression for the individual, where evidence of progression is defined as progression of non-target lesions, the appearance of new lesions or as determined by an investigator, impute a change from baseline at week 6 as 20%. If the patient has an imputed value from (a), use the maximum of 20% or the imputed value in the tumour size.

(c) If there is no evidence of progression for the individual, use the imputed value calculated in (a) if data available. If no data are available, assume that the data are missing completely at random, the patient will be excluded from the analysis.

(d) If it is known that the patient has died, impute a change from baseline at week 6 as the maximum of 20% or the largest percentage increase calculated from actual or imputed data.

The best change in target lesion (TL) tumour size is the biggest decrease or smallest increase in tumour size from baseline prior to progressive disease (PD). The percentage change from baseline with be calculated by dividing the best change from baseline by the sum of the TLs at baseline and multiplying by 100 (ie, (best change from baseline) / baseline * 100). Any visit responses where the RECIST TL is not evaluable (NE) for any reason will be excluded prior to calculation of best change in tumour size as these measurements are judged to be unreliable.

Consistent with the approach taken for change in tumour at week 6, for each individual patient with no post baseline assessments the following imputation methods will be applied:

(a) If there is evidence of progression or it is known that the patient has died, the percentage change from baseline will be imputed as 20%.

(b) If there is no evidence of progression, assume that the data is missing completely at random, the patient will be excluded from the analysis.

4.2.5 Duration of response

The median duration of response based on the investigator's assessment of RECIST 1.1 will be summarised with corresponding 95% CIs split by treatment arm using K-M techniques. Only patients who had a response will be included in these summary tables. Formal statistical

testing between treatment groups will not be performed. In addition median time to onset of response from randomisation will be reported based on summary statistics.

4.2.6 RECIST 1.1 data by BICR

RECIST data from the BICR will be listed for all patients selected for BICR review.

Analyses will be conducted to identify any potential ascertainment bias in the investigator assessments compared with BICR according to RECIST 1.1. See section 4.2.2.1 c for full details.

If any potential ascertainment bias is identified, a BICR of additional patient scans may be performed.

4.2.7 Patient reported outcomes (PRO)

4.2.7.1 Lung cancer symptom scale (LCSS)

(a) Time to symptom progression:

Time to symptom progression (ASBI) will be analysed as described for the primary analysis of PFS. However subgroup analyses, treatment interaction testing and sensitivity analyses will not be performed (with the exception of attrition bias).

Attrition bias will be assessed by repeating the ASBI analysis except that the actual clinically meaningful symptom deterioration event times, rather than the censored times, of patients who died in the absence of clinically meaningful symptom deterioration immediately following two, or more, non-evaluable LCSS assessments will be included. In addition, patients who take subsequent therapy prior to deterioration or death will be censored at their last evaluable LCSS assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring where the censoring indicator of the ASBI analysis is reversed.

In addition, time to symptom progression for each of the 6 individual symptoms (appetite, fatigue, coughing, shortness of breath, blood in sputum and pain) will be compared between treatment groups using a stratified log-rank test as described for the primary analysis of PFS. The HR and 95% CI for each symptom will be presented graphically on a forest plot. P-values will not be calculated for these supportive analyses.

In addition, two sensitivity analyses will be performed for time to symptom progression (ASBI) only: 1) where RECIST 1.1 progression is considered a symptom progression event; and 2) where patients are censored at RECIST 1.1 progression provided symptom progression has not yet occurred. Analyses will be performed using a stratified log-rank test as for the primary analysis of symptom progression (ASBI), attrition bias will not be assessed for these sensitivity analyses.

(b) Symptom improvement rate:

A summary of symptom improvement rate (ASBI) will be produced. Symptom improvement rate will be analysed as described for the analysis of ORR.

Supportive analyses will be performed for the individual symptoms from the ASBI. The symptom improvement rate for each of the 6 individual symptoms will be compared between treatment groups using a logistic regression model as described for ORR. The odds ratio and 95% CI for each symptom will be presented graphically on a forest plot. If there are too few patients with improvements available for a meaningful analysis of a particular symptom (it is not considered appropriate to present analyses where there are less than 10 patients in either treatment arm with improvements in a symptom), the relationship between that symptom and improvement rate will not be formally analysed. In this case, only descriptive summaries will be provided.

(c) LCSS score summaries

The individual item scores from the LCSS for symptom distress and interference with activity levels will be summarized using descriptive statistics only (formal statistical testing between treatment groups will not be performed). Absolute values and changes from baseline will also be summarised over time for each treatment group for the ASBI and for each of the 6 individual items that comprise the ASBI. In addition, response rates for each treatment group will be summarized for each item.

A summary of compliance rate and evaluability rate will be provided, by assessment time point and also for overall.

4.2.7.2 Short-Form Health Survey (SF-36v2)

The scores for each of the 8 health domain scales and for each of the physical and mental component summary measures will be summarised by absolute values and changes from baseline at each post-baseline assessment. There will not be any formal statistical testing.

4.2.8 Safety

The following sections describe the planned safety summaries for AEs, vital signs, laboratory parameters, echocardiograms, WHO performance status and visual assessments. However, additional safety tables (not specified in this SAP) may need to be produced to aid interpretation of the safety data.

4.2.8.1 Adverse events

AEs, including MedDRA preferred term and maximum CTCAE grade, will be listed individually by patient and treatment group. Changes in CTC grade for individual adverse events (e.g., dates and duration at each grade) will also be listed.

Events will be defined as treatment emergent if they onset, or worsen (by investigator report of a change in intensity), during the treatment period (by summarising based on date of onset). If partial dates of onset occur, a conservative approach will be followed and events will be assumed to be treatment emergent unless there is convincing evidence to the contrary.

AEs and the number of patients experiencing the AEs will be summarised and will include but not be limited to:

All AEs

- AEs summary by AE category (refers to topline summary of the categories listed below)
- AEs by system organ class (SOC) and preferred term (PT)
- AEs by PT with frequency ≥ 5% in any treatment group
- AEs by SOC and PT, by maximum CTCAE grade

Serious adverse events (SAEs)

- SAEs by SOC and PT
- SAEs by PT with frequency ≥ 2% in any treatment group
- SAEs with serious criterion hospitalisation by SOC and PT
- SAEs with serious criterion hospitalisation by PT with frequency ≥ 2% in any treatment group
- Listing of key information for SAEs
- Listing of SAEs in screening failure patients related to biopsy

Death

- All deaths
- Listing of deaths
- AEs with outcome of death by SOC and PT
- Listing of key information of AEs leading to death

Adverse events leading to discontinuation of treatment or dose interruptions

- AEs leading to selumetinib/placebo dose modification, by SOC and PT. Note, dose modification include dose reduction or interruption.
- AEs leading to selumetinib/placebo dose reduction, by SOC and PT
- AEs leading to selumetinib/placebo dose interruption, by SOC and PT
- AEs leading to docetaxel dose modification, by SOC and PT

- AEs leading to docetaxel dose reduction, by SOC and PT
- AEs leading to docetaxel dose interruption, by SOC and PT
- AEs leading to discontinuation of selumetinib/placebo and docetaxel by SOC and PT
- AEs leading to discontinuation of selumetinib/placebo by SOC and PT
- AEs leading to discontinuation of docetaxel, by SOC and PT
- Listing of key information of AEs leading to discontinuation of selumetinib/placebo
- Listing of key information of AEs leading to discontinuation of docetaxel

Other

- Adverse events of special interest (list of groups and specific PTs specified in Appendix E)
- AEs of special interest by grouped AE
- Listing of key information for adverse events of special interest
- AEs of CTCAE grade 3 or higher by SOC and PT
- AEs of CTCAE grade 3 or higher by PT with frequency ≥ 2% in any treatment group
- Duration of adverse events of special interest by CTC grade by treatment group. For each specified event of special interest, information will be summarised as follows:
  - Number of patients with event (n, %); Number of episodes of event (n); Median total duration of event (days, range); Median duration of CTC grade 1 (days, range), Median duration of CTC grade 2 (days, range); Median duration of CTC grade 3 (days, range); Median duration of CTC grade 4 (days, range); Median duration of CTC grade 3 or 4 (days, range)
- Other significant AEs (ICH E3 definition to be applied at blind data review)
- Listing of key information of Other significant AEs

When applying a cut-off (i.e., x %), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency of 9.9% will not appear if a cut-off is 10%).

Prevalence plots will be presented for grouped AEs of special interest (key interest only).

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

4.2.8.2 Laboratory assessments

All laboratory data and changes from baseline will be listed. Flags will be applied to values falling outside - reference ranges (which will be explicitly noted on these listings where applicable), and to values for which CTC grading applies.

Absolute values and change from baseline will be summarised using descriptive statistics at each scheduled time point (unscheduled data are included by windowing visits) by treatment group. Descriptive statistics will include number of non-missing patients (n), mean, standard deviation, median, first and third quartiles (Q1 and Q3), minimum and maximum.

Data summaries and listings will be provided in International System (SI) of units.

For laboratory parameters for which CTC grading exists, shift tables by treatment group for laboratory values by worst CTC grade during treatment will be produced. In addition, the number of patients with a 0-, 1-, 2-, 3- or 4-grade shift from baseline will be summarised for each of these parameters:

- Haematology: Haemoglobin – hypo, Lymphocytes, absolute count – hypo, Neutrophils, absolute count – hypo, Platelets – hypo
- Clinical chemistry – liver biochemistry: Alanine Aminotransferase (ALT) – hyper, AST – hyper, Alkaline Phosphatase (ALP) – hyper, Total bilirubin – hyper; Albumin – hypo
- Clinical chemistry – electrolytes: Magnesium – hypo and – hyper, Phosphate – hypo; Sodium – hypo and – hyper; Potassium – hypo and – hyper; Corrected calcium – hypo and – hyper
- Clinical chemistry – renal biochemistry: Creatinine – hyper

A summary of treatment-emergent laboratory variable changes outside the predefined criteria at any time during treatment will be produced for parameters that are not classified by CTC grading (refer section 3.5.2).

Scatter plots (shift plots) of baseline to maximum value on treatment will be produced for: ALT, AST, ALP, total bilirubin, creatinine, and urea.

Scatter plots (shift plots) of baseline to minimum value on treatment will be produced for: haemoglobin, lymphocyte count, absolute; neutrophils count, absolute; platelet count; and albumin.

Box-plots of absolute values by study day, and box-plots of change from baseline by study day, will be presented for haemoglobin; neutrophil count, absolute; lymphocyte count, absolute; platelet count; AST; ALT; ALP; Total bilirubin; albumin; creatinine and urea.

Hy's law

Following summaries by treatment group will include number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study
  - ALT ≥ 3x –< 5x, ≥ 5x – <10x and ≥ 10x Upper Limit of Normal (ULN) during the study
  - AST ≥ 3x – <5x, ≥ 5x – <10x and ≥ 10x ULN during the study
  - Total bilirubin ≥2x ULN during the study
- Narratives will be provided for patients who have ALT ≥ 3x ULN plus Total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus Total bilirubin ≥ 2x ULN at any visit.

Individual patient data where ALT or AST plus Total bilirubin are elevated at any time will be listed and presented in a plot.

Plots of ALT and AST vs. Total bilirubin by treatment group will also be produced with reference lines at 3×ULN for ALT, AST, and 2×ULN for Total bilirubin. In each plot, Total bilirubin will be in the vertical axis.

Additional summaries will include a shift table for urinalysis comparing baseline value to maximum value during study treatment.

4.2.8.3 Neutropenia and infection

Information from INFNEUT eCRF will be summarised under the Table title "Overview of grade 3 or 4 neutropenia" by number and percentage of patients, and by number of episodes, by treatment group with

- Grade 3 or 4 neutropenia [patients n (%); episodes n]
- Grade 3 or 4 neutropenia with recovery before next cycle [patients n (%); episodes n]
- Grade 3 or 4 neutropenia with resulting infection [patients n (%); episodes n]

Information from INFNEUT eCRF will be also summarised by treatment cycle under the Table title "Summary of grade 3 or 4 neutropenia by treatment cycle" by treatment group, by number and percentage of patients, and then by number of episodes:

- Grade 3 or 4 neutropenia at any time, Pre-study treatment, During combination treatment (subset by Cycle 1, Cycle 2 etc.), During selumetinib/placebo monotherapy[1], During docetaxel monotherapy[2], After discontinuation of all study treatment

Information on the earliest episode of grade 3 or 4 neutropenia from INFNEUT eCRF be also summarised by treatment cycle under the Table title "Summary of first episode grade 3 or 4 neutropenia by treatment cycle" by treatment group, and number and proportion of patients:

- Grade 3 or 4 neutropenia at any time, Pre-study treatment, During combination treatment (subset by Cycle 1, Cycle 2 etc.), During selumetinib/placebo monotherapy, During docetaxel monotherapy, After discontinuation of all study treatment

Information from INFNEUT eCRF on infection resulting from grade 3 or 4 neutropenia will be summarised by preferred term under the Table title "Summary of infection resulting from grade 3 or 4 neutropenia by preferred term" by treatment group and number and percentage of patients with:

- Infection as a result of grade 3 or 4 neutropenia, PT1, PT2 (ordered in decreasing order of frequency)

- Highest body temperature reported during Neutropenia of grade 3/4

A listing of key information for patients with grade 3 or 4 neutropenia will be produced.

4.2.8.4 Infection diagnostic investigations

A summary table of information on diagnostic investigations for infection (INFDI eCRF) will include the number and percentage of patients by treatment with

- AE of infection (AE in infection SOC recorded on AE eCRF)

- Diagnostic investigation for infection (with subsets for Microscopy, culture and sensitivity, Serological tests, Nucleic acid-based tests, X-ray and Other)

---

[1] This category will be for events occurring in patients continuing to receive Selumetinib/placebo after permanent discontinuation of docetaxel

[2] This category will be for events occurring in patients continuing to receive docetaxel after permanent discontinuation of Selumetinib/placebo

Free text from the "Results" field will be included in listings and not summarised.

4.2.8.5 Vital signs

Vital signs (pulse rate, resting systolic and diastolic blood pressure and weight) will be summarised over time in terms of absolute values and changes from baseline at each scheduled time point by treatment group using descriptive statistics. Descriptive statistics will include number of non-missing patients (n), mean, standard deviation, median, first and third quartiles (Q1 and Q3), minimum and maximum.

Vital signs (pulse rate, systolic blood pressure, diastolic blood pressure, weight and height [baseline only]) will be listed. Flags will be applied to values falling outside - reference ranges (which will be explicitly noted on these listings where applicable), and to values for which CTC grading applies (e.g., CTC grading for Hypertension (numeric criteria) flags to be applied to listings for systolic and diastolic blood pressure).

Box-plots of absolute values by study day by treatment group, and change from baseline by study day by treatment group will also be produced for systolic blood pressure and diastolic blood pressure.

4.2.8.6 ECG

All ECG data will be listed.

4.2.8.7 Echocardiogram

All echocardiogram parameters will be listed.

The LVEF and change from baseline in LVEF will be summarised by treatment group using standard descriptive summary statistics at each scheduled time point (unscheduled data are included by windowing visits).

Box-plots of absolute values by study day by treatment group, and change from baseline by study day by treatment group will be produced for LVEF. A shift plot showing LVEF at baseline vs minimum on-treatment value will also be produced.

Given the known variability in LVEF assessments, a reduction from baseline in LVEF of ≥10 percentage points will be considered a real change in LVEF. An LVEF value of at least 55% is a prerequisite for eligibility therefore this is the minimum baseline value assumed in this study.

A summary table of changes in LVEF by treatment group will include the number and proportion of patients who have:

- Baseline LVEF assessment.
- Post-baseline (at any time point following dosing) LVEF assessment

- LVEF decrease of ≥ 10 percentage points (at any time point following dosing)
- Absolute LVEF < 55% (at any time point following dosing)
- LVEF decrease of ≥ 10 percentage points and to < 55% (occurring at the same echocardiography assessment, at any time point following dosing)

4.2.8.8 Physical examination

All individual physical examination data will be listed.

4.2.8.9 Other safety data

Data from positive pregnancy tests and troponin I measurements will be listed and not summarised.

Visual assessments (including optical coherence tomography) will be listed.

4.2.9 WHO performance status

WHO performance status will be summarised through looking at a shift table of baseline score against worst score post-baseline for FAS and listed.

4.2.10 PK data

Plasma selumetinib and N-desmethyl selumetinib plasma concentrations will be listed for each patient per dose and dosing day but will not be summarised.

4.2.11 Demographic data and baseline data

The following baseline characteristics will be listed for each patient and summarised for all patients in the FAS (unless otherwise specified) by treatment group:

- Patient disposition for all prescreened/screened patients
- Inclusion in analysis populations
- Important protocol deviations
- *KRAS* mutation subtype for all patients and at baseline for the FAS
- Demography (age (for the FAS calculated as age at randomisation and for the All Patients at main informed consent if available or pre-screening informed consent if not), age (as above) group [< 65, ≥ 65 years], sex, race and ethnic group). These will also be presented for all patients and by *KRAS* mutation subtype.
- Patient characteristics including height and weight
- Patient recruitment by country and centre

- Previous disease-related treatment modalities
- Number of regimens of previous chemotherapy at baseline
- Previous disease-related chemotherapy treatments
- Past and current medical history
- Relevant surgical history
- Disease characteristics at baseline. These will also be presented for all patients and by *KRAS* mutation subtype.
- Extent of disease at baseline
- Primary tumor and TNM classification at original diagnosis
- Time from most recent disease progression to randomisation
- Allowed and disallowed concomitant medications (presented by ATC classification system) for all patients in the safety analysis set
- Nicotine use, categorised (never, current, former)
- Stratification factors as per IVRS and eCRF data
- Post-discontinuation disease-related anticancer therapy for all patients in the safety analysis set

4.2.12 Exposure

Exposure will be summarised for safety analysis set.

The following summaries related to treatment and dose reductions/interruptions/delays will be produced:

- Summary of duration of exposure of selumetinib/placebo.
- Summary of total number of cycles of docetaxel received
- PID and RDI of selumetinib/placebo
- PID and RDI of docetaxel
- Summary of time to first dose interruption, reduction or discontinuation to selumetinib/placebo

- Summary of interruptions and reductions of selumetinib/placebo and docetaxel by cycles (up to cycle 6)

- Summary of delays and reductions of docetaxel overall and by cycles (up to cycle 6)

Summaries of routine prophylaxis G-CSF administration by docetaxel cycle and the number of patients receiving G-CSF and the reasons for administration will also be produced.

All treatment information data will be listed by treatment group.

5. INDEPENDENT DATA MONITORING COMMITTEE (IDMC)

This study will use an external IDMC to perform ongoing safety analyses. The IDMC guidelines document will present the details and timings of each of the reviews.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AZ, and do not have any major conflict of interest.

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

The IDMC guidelines document will also contain details of the IDMC members and clearly define the responsibilities of the IDMC.

6. REFERENCES

Anderson et al 2004
Anderson P.K, Hansen, M.G., Klein, J.P. Regression analysis of restricted mean survival time based on Pseudo-observations. Lifetime Data Anal. 2004 Dec:10(4): 335-350

Berry et al.1991
Berry G, Kitchin RM, Mock PA. A comparison of two simple hazard ratio estimators based o n the logrank test Statistics in Medicine 1991; 10:749-755

Brazier et al.2002
Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002 Mar;21(2):271-92

Breslow, 1974
Breslow, NE. Covariance Analysis of Censored Survival Data. Biometrics 1974; 30:89–99.

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

Collett, 2003
Collett D. Modelling Survival Data in Medical Research. 2nd ed. Boca Raton (FL): Chapman & Hall/CRC; 2003.

De Marinis et al 2008
De Marinis F, Pereira JR, Fossella F, Perry MC, Reck M, Salzberg M, et al. Lung Cancer Symptom Scale outcomes in relation to standard efficacy measures: an analysis of the phase III study of pemetrexed versus docetaxel in advanced non-small cell lung cancer. J Thorac Oncol 2008;3(1):30-36.

Dodd et al 2011
Dodd LE, Korn EL, Freidlin B, Gray R, and Battacharya S. An audit strategy for progression-free survival. Biometrics 2011;67(3):1092-1099.

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

Glimm et al 2009
Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med 2009: 29:219-228

Hollen and Gralla 2000
Hollen P, Gralla R. Clinical vs. statistical significance: using the LCSS Quality of Life Instrument and Karnofsky Performance Status (KPS) to approach the problem in patients with non-small cell lung cancer (NSCLC). Proc Am Soc Clin Oncol. 2000;19(abstr 2510).

Lan and DeMets 1983
Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70;3:659–63.

Lee S et al, 2015
Lee S, Cao J, May T, Li J, Corona L, et al. (2015) Performance Characteristics of a PCR Assay for the Detection of KRAS Mutations in Formalin-Fixed Paraffin-Embedded Tissue Samples of Non-Small Cell Lung Cancer. J Mol Biomark Diagn 6:246.

McHorney et al, 1993
McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993 Mar;31(3):247-63

McHorney et al, 1994
McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994 Jan;32(1):40-66.

Rowen et al 2009
Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health Qual Life Outcomes. 2009 Mar 31;7:27

Royston and Parmar 2011
Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Stat Med 2011 30:2409-2421

Sarna et al 2008
Sarna L, Cooley ME, Brown JK, Chernecky C, Elashoff D, Kotlerman J. Symptom severity 1 to 4 months after thoracotomy for lung cancer. Am J Crit Care. 2008;17(5):455-467; quiz 468.

Sellke and Siegmund 1983
Sellke T and Siegmund D. Sequential analysis of the proportional hazards model. Biometrika 1983;70: 315–326

Stone et al. 2015
Stone A, Macpherson E, Smith A, and Jennison C. Model free audit methodology for bias evaluation of tumour progression in oncology. PharmStats (submitted)

Sun and Chen 2010
Sun X and Chen C. Comparison of Finkelstein's Method With the Conventional Approach for Interval-Censored Data Analysis. Stat Biopharm Research 2010;2:1:97-108.

Ware et al 1993
Ware JE, Snow K, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. 1993

Ware et al, 2007
Ware JE, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME (2007). User's manual for the SF-36v2 Health Survey (2nd ed). Lincoln, RI : QualityMetric Incorporated.

7. APPENDIX

Appendix A Laboratory parameters – Minimum/Maximum of interest

Table 8 Laboratory parameters - Minimum/Maximum of interest

Laboratory Assessment	Maximum of interest	Minimum of interest
ALT	Yes	
AST	Yes	
ALP	Yes	
Total bilirubin	Yes	
Corrected calcium [a]	Yes	Yes
Magnesium	Yes	Yes
Sodium	Yes	Yes
Potassium	Yes	Yes
Creatinine	Yes	
Urea	Yes	
Haemoglobin		Yes
Lymphocyte count		Yes
Absolute neutrophils count		Yes
Absolute platelet count		Yes
Albumin		Yes
Total protein		Yes

[a] Corrected calcium will be calculated using the following formula: Corrected calcium = Total Calcium (mmol/L) + ([40 – Albumin (G/L)] x 0.02)

Appendix B BICR Audit Plan

1. Introduction

At the final analysis, potential ascertainment bias in the investigator determination of PFS will be assessed using the BICR audit strategies proposed by Dodd et al 2011 and Stone et al. 2015. For both methods, a sample of 220 patients will be randomly selected as the audit sample. The audit size is determined by ensuring sufficient power for the inference about PFS effect in Dodd method and high specificity for confirming consistency between BICR and LE in the model-free method in Stone et al. 2015. The details are presented below within each method.

2. Audit plan based on Dodd method

Let $UCB_{S,\alpha}$ denote the one-sided (1-α)% upper confidence bound as defined in Dodd et al 2011 based on the audit sample, and γ denote the clinical irrelevant factor. If $UCB_{S,\alpha} > \gamma$, then a full BICR will be conducted, where α is set to be 2.5% and γ is set to be log(1).

Derivation:

Let θ denote the log-hazard ratio in the underlying population. Let S refer to the audit sample, $\bar{S}$ refer to the unaudited subset, and F refer to the trial population.

Dodd (2011) proposed a more efficient estimator for θ than the naïve estimator $\log(\widehat{HR})_{BICR,S}$. The estimator is obtained as:

$$\hat{\theta} = \log(\widehat{HR})_{BICR,S} + \hat{\rho}\sqrt{\delta}(1-\delta)\sqrt{\frac{\hat{V}_{BICR,\bar{S}}}{\hat{V}_{LE,F}}}(\log(\widehat{HR})_{LE,\bar{S}} - \log(\widehat{HR})_{LE,S})$$

where V stands for variance, δ is the proportion of study subjects audited, and ρ is the correlation between $\log(\widehat{HR})_{BICR,S}$ and $\log(\widehat{HR})_{LE,S}$. With a variance estimator $\hat{V} = \hat{V}_{BICR,S}(1-\hat{\rho}^2(1-\delta))$ and normal approximation, the one-sided (1-α)% upper confidence bound for θ is $\hat{\theta} + Z_{1-\alpha}\sqrt{\hat{V}}$. A conclusion of BICR audit being sufficient is drawn if $\hat{\theta} \pm Z_{1-\alpha}\sqrt{\hat{V}} < \gamma$, where γ is the clinical irrelevant factor.

3. Audit plan based on model-free method

Let $\widehat{HRR}_F = \frac{\widehat{HR}_{BICR,F}}{\widehat{HR}_{LE,F}}$ denote the hazard ratio ratio (HRR) estimated from all patients in the study study and let $\widehat{HRR}_S = \frac{\widehat{HR}_{BICR,S}}{\widehat{HR}_{LE,S}}$ denote that from the audit sample. The model-free method tests the hypothesis that this HRR estimate equals or exceeds a pre-specified threshold HRR_U, i.e., $H_0: \log(\widehat{HRR}_F) \geq \log(HRR_U)$ vs. $H_A: \log(\widehat{HRR}_F) < \log(HRR_U)$, based on the audit sample.

The model-free method makes inference about the estimated HRR from the trial (i.e., $\widehat{HRR}_F$); therefore, hypothesis testing is needed only in the auditing stage because $\widehat{HRR}_F$ will be known

exactly when a full BICR is conducted. If the null hypothesis is rejected based on BICR audit, there is enough evidence in the audit sample to conclude consistency between BICR and LE; otherwise, the $\widehat{HRR}_F$ will be obtained via full BICR, and can be compared with HRR_U directly without hypothesis testing.

As suggested in Stone et al. 2015, α is set to be 10% (one-sided) to give 90% sensitivity in detecting inconsistency between BICR and LE, and HRR_U is set to be 1.25.

Derivation:

The conditional distribution of $\log(\widehat{HRR})_S$ given $\log(\widehat{HRR})_F = \log(\widehat{HRR})_F$ is $\log(\widehat{HRR})_S | \log(\widehat{HRR})_F = \log(\widehat{HRR})_F \sim N(\log(\widehat{HRR})_F, I_S^{-1} - I_F^{-1})$. To test the hypothesis H_0 as described above, a confidence bound can be constructed and the null hypothesis can be rejected if $\log(\widehat{HRR})_S < \log(HRR_U) - Z_{1-\alpha}\sqrt{\hat{I}_S^{-1} - \hat{I}_F^{-1}}$. For a 1:1 randomization, $\hat{I}_F^{-1} = Var\left(\log(\widehat{HRR})_F\right) = Var\left(\log(\widehat{HR}_{BICR,F}) - \log(\widehat{HR}_{LE,F})\right) = \frac{4}{D_{BICR,F}} + \frac{4}{D_{LE,F}} - 8\hat{\rho}\sqrt{\frac{1}{D_{BICR,F}D_{LE,F}}}$, where ρ is the correlation between $\log(\widehat{HR})_{BICR,F}$ and $\log(\widehat{HR})_{LE,F}$. Assume $\frac{D_{BICR,F}}{D_{LE,F}} = \varphi$, then $I_F^{-1} = \frac{4}{D_{LE,F}}\frac{1+\varphi-2\hat{\rho}\sqrt{\varphi}}{\varphi}$. Similarly, $I_S^{-1} = \frac{4}{D_{LE,S}}\frac{1+\varphi-2\hat{\rho}\sqrt{\varphi}}{\varphi}$, assuming $\frac{D_{BICR,S}}{D_{LE,S}} = \varphi$. If the null hypothesis is rejected, i.e., $\log(\widehat{HRR})_S < \log(HRR_U) - Z_{1-\alpha}\sqrt{\hat{I}_S^{-1} - \hat{I}_F^{-1}}$, then a conclusion of BICR and LE giving consistent HR estimates can be drawn based on the audit sample.

Appendix C AstraZeneca project-specific normal range for laboratory parameters

Table 9 AstraZeneca project-specific normal range for laboratory parameters

Test	Gender	SI Unit	LLN	ULN
Alanine aminotransferase (ALT)	M and F	U/L	0	50
Aspartate aminotransferase (AST)	M and F	U/L	0	45
Alkaline phosphatase	M and F	U/L	20	130
Total bilirubin	M and F	µmol/L	2	21
Corrected calcium	M and F	mmol/L	2.30	2.74
Calcium – phosphate product	M and F	mmol2/L2	-	4.5
Magnesium	M and F	mmol/L	0.74	1.48
Sodium	M and F	mmol/L	136	145
Potassium	M and F	mmol/L	3.5	4.9
Creatinine	M and F	µmol/L	53	106
Creatine Phosphokinase	M	U/L	55	170
Creatine Phosphokinase	F	U/L	30	135
Urea nitrogen	M and F	mmol/L	2.5	6.7
Haemoglobin	M	g/L	135	180
Haemoglobin	F	g/L	120	160
Lymphocyte count	M and F	10**9/L	1	4.8
Absolute neutrophils count	M and F	10**9/L	1.8	7.8
Monocyte count	M and F	10**9/L	0	0.8
Leukocyte count	M and F	10**9/L	4.4	11.3
Eosinophil count	M and F	10**9/L	0	0.45
Basophil count	M and F	10**9/L	0	0.2
Absolute platelet count	M and F	10**9/L	150	450
Phosphate	M and F	µmol/L	1.12	1.45
Albumin	M and F	g/L	32	56
Prothrombin Intl. Normalized Ratio	M and F		0.8	1.4
Troponin I	M and F		0	0.2
Total protein	M and F	g/L	60	80

F: female; M: male

Appendix D NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 grades for laboratory parameters

Table 10 CTCAE grades for laboratory parameters

Lab Parameter	NCI Common Terminology Criteria for Adverse Events (CTCAE) v4.0			
	Grade 1	Grade 2	Grade 3	Grade 4
Haematology				
White blood cell decreased	< LLN – 3.0 x 10**9/L	< 3.0 – 2.0 x 10**9/L	< 2.0 – 1.0 x 10**9/L	< 1.0 x 10**9/L
Absolute neutrophil count decreased	< LLN – 1.5 x 10**9/L	< 1.5 – 1.0 x 10**9/L	< 1.0 – 0.5 x 10**9/L	< 0.5 x10**9/L
Absolute lymphocyte count decreased	< LLN – 0.8 x 10**9/L	< 0.8 – 0.5 x 10**9/L	< 0.5 – 0.2 x 10**9/L	< 0.2 x 10**9/L
Haemoglobin increased	> ULN - (ULN + 200) g/L	> (ULN + 200) – (ULN + 400) g/L	> (ULN + 400) g/L	Not applicable
Haemoglobin decreased (Anaemia)	< LLN – 100 g/L	< 100 – 80 g/L	< 80 g/L	Not applicable
Platelets decreased	< LLN – 75.0 x 10**9/L	< 75.0 – 50.0 x 10**9/L	< 50.0 – 25.0 x 10**9/L	< 25.0 x 10**9/L
Albumin decreased	< LLN – 30 g/L	< 30 – 20 g/L	< 20 g/L	Not applicable
Clinical Chemistry				
ALT increased	> ULN – 2.5 x ULN	> 2.5 -5.0 x ULN	> 5.0 -20.0 x ULN	> 20.0 x ULN
ALP increased	> ULN - 2.5 x ULN	> 2.5 - 5.0 x ULN	> 5.0 - 20.0 x ULN	> 20.0 x ULN
AST increased	> ULN – 2.5 x ULN	> 2.5 -5.0 x ULN	> 5.0 -20.0 x ULN	> 20.0 x ULN
Blood bilirubin increased	> ULN - 1.5 x ULN	> 1.5 - 3.0 x ULN	> 3.0 - 10.0 x ULN	> 10.0 x ULN

Table 10 CTCAE grades for laboratory parameters

Lab Parameter	NCI Common Terminology Criteria for Adverse Events (CTCAE) v4.0			
	Grade 1	**Grade 2**	**Grade 3**	**Grade 4**
Calcium low[a] (Hypercalcaemia)	Corrected serum calcium of < LLN - 2.0 mmol/L	Corrected serum calcium of < 2.0 - 1.75 mmol/L	Corrected serum calcium of < 1.75 - 1.5 mmol/L	Corrected serum calcium of < 1.5 mmol/L
Calcium high[a] (Hypercalcaemia)	Corrected serum calcium of > ULN - 2.9 mmol/L	Corrected serum calcium of > 2.9 - 3.1 mmol/L	Corrected serum calcium of > 3.1 - 3.4 mmol/L	Corrected serum calcium of > 3.4 mmol/L
Creatinine increased	> ULN - 1.5 x ULN	> 1.5 - 3.0 x ULN	> 3.0 - 6.0 x ULN	> 6.0 x ULN
Creatine Phosphokinase increased	> ULN – 2.5 x ULN	> 2.5 – 5.0 x ULN	> 5.0 – 10.0 x ULN	> 10.0 x ULN
Magnesium high (Hypermagnesaemia)	> ULN - 1.23 mmol/L	Not applicable	> 1.23 - 3.30 mmol/L	> 3.30 mmol/L
Magnesium low (Hypomagnesaemia)	< LLN - 0.5 mmol/L	< 0.5 - 0.4 mmol/L	< 0.4 - 0.3 mmol/L	< 0.3 mmol/L
Potassium high (Hyperkalemia)	> ULN - 5.5 mmol/L	> 5.5 - 6.0 mmol/L	> 6.0 - 7.0 mmol/L	> 7.0 mmol/L
Potassium low (Hypokalemia)	< LLN - 3.0 mmol/L	Not applicable	< 3.0 - 2.5 mmol/L	< 2.5 mmol/L
Sodium high (Hypernatremia)	> ULN – 150 mmol/L	> 150 – 155 mmol/L	> 155 – 160 mmol/L	> 160 mmol/L
Sodium low (Hyponatremia)	< LLN – 130 mmol/L	Not applicable	< 130 – 120 mmol/L	< 120 mmol/L

Because many institutions have differences for normal ranges of metabolic, laboratory, and haematology values, the CTCAE often uses the terms ULN and LLN in lieu of actual numerical values. In some cases, an institution's LLN might be beyond the range specified for a Grade 1. In this case, the institutional limits of normal should take precedence over the CTCAE values.

[a] Corrected Calcium (mmol/L) = Total Calcium (mmol/L) + ([40 – Albumin (G/L)] x 0.02), ALP Alkaline phosphatase, ALT Alanine aminotransferase, AST Aspartate aminotransferase, LLN lower limit of normal, NCI National Cancer Institute, ULN upper limit of normal

Appendix E Adverse Events of Special Interest (AESIs) and Grouped Terms

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
Effects on the skin and mucous membranes	Rash acneiform	Acne
		Acne pustular
		Dermatitis acneiform
		Folliculitis
		Rash
		Rash follicular
		Rash papular
		Rash pustular
Effects on the skin and mucous membranes	Rash, non-acneiform	Drug eruption
		Dermatitis
		Dermatitis exfoliative
		Erythema
		Exanthema
		Exfoliative rash
		Rash generalised
		Rash maculo-papular
		Rash macular
		Rash maculovesicular
		Rash vesicular
		Rash erythematous
		Rash pruritic
		Rash morbilliform
		Skin erosion
		Skin exfoliation
Effects on the skin and mucous membranes	Rashes	Acne
		Acne pustular
		Dermatitis acneiform
		Folliculitis
		Rash

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Rash follicular
		Rash papular
		Rash pustular
		Drug eruption
		Dermatitis
		Dermatitis exfoliative
		Erythema
		Exanthema
		Exfoliative rash
		Rash generalised
		Rash maculo-papular
		Rash macular
		Rash maculovesicular
		Rash vesicular
		Rash erythematous
		Rash pruritic
		Rash morbilliform
		Skin erosion
		Skin exfoliation
Effects on the skin and mucous membranes	Dry skin effects	Dry skin
		Eczema
		Skin fissures
		Xeroderma
		Xerosis
Effects on the skin and mucous membranes	Paronychia effects	Nail bed disorder
		Nail bed infection
		Nail bed inflammation
		Nail bed tenderness
		Onycholysis
		Paronychia
Effects on the skin and	Oral mucositis effects	Aphthous ulcer

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
mucous membranes		Gingivitis
		Mouth ulceration
		Oral discomfort
		Oral mucosa erosion
		Oral mucosal eruption
		Oral pain
		Stomatitis
Effects on the eyes	Retinal vein occlusion effects	Retinal haemorrhage
		Retinal vascular occlusion
		Retinal vascular thrombosis
		Retinopathy haemorrhagic
		Retinal vein occlusion
		Retinal vein thrombosis
		Venous stasis retinopathy
		Retinal vascular disorder
Effects on the eyes	RPED / CSR effects	Acquired pigmented retinopathy
		Chorioretinopathy
		Detachment of macular retinal pigment epithelium
		Detachment of retinal pigment epithelium
		Maculopathy
		Retinal pigment epithelial tear
		Retinal pigment epitheliopathy
		Retinopathy
		Subretinal fluid
Effects on the eyes	Other retinal effects	Blindness
		Blindness unilateral
		Cystoid macular oedema
		Exudative retinopathy
		Eye disorder
		Fundoscopy abnormal
		Macular cyst

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Macular oedema
		Macular ischaemia
		Optical coherence tomography abnormal
		Paraneoplastic retinopathy
		Retinal aneurysm
		Retinal cyst
		Retinal detachment
		Retinal disorder
		Retinal exudates
		Retinal function test abnormal
		Retinal infarction
		Retinal ischaemia
		Retinal neovascularisation
		Retinal oedema
		Retinal tear
		Retinal vascular disorder
		Retinopathy proliferative
		Retinal toxicity
		Subretinal haematoma
		Visual acuity reduced
		Visual acuity tests abnormal
		Visual field defect
		Visual field tests abnormal
		Visual impairment
		Vision blurred
		Vitreous disorder
		Vitreous floaters
Effects on the lungs	Interstitial lung disease (ILD-type) effects	Acute interstitial pneumonitis
		Alveolar lung disease
		Alveolitis
		Alveolitis necrotising

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Diffuse alveolar damage
		Idiopathic pulmonary fibrosis
		Interstitial lung disease
		Lung infiltration
		Pneumonitis
		Progressive massive fibrosis
		Pulmonary toxicity
Effects on the lungs	Dyspnoea effects	Dyspnoea
		Dyspnoea at rest
		Dyspnoea exertional
Effects on blood cells	Febrile neutropenia effects	Febrile neutropenia
		Neutropenic infection
		Neutropenic sepsis
Effects on blood cells	Neutropenia effects	Cyclic neutropenia
		Leucopenia
		Neutropenia
		Neutrophil count decreased
		White blood cell count decreased
Effects on blood cells	Thrombocytopenia effects	Megakaryocytes decreased
		Platelet count decreased
		Platelet maturation arrest
		Platelet production decreased
		Platelet toxicity
		Thrombocytopenia
Effects on blood cells	Erythropenia effects	Anaemia
		Anaemia macrocytic
		Aplasia pure red cell
		Aplastic anaemia
		Erythroblast count decreased
		Erythroid maturation arrest
		Erythropenia

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Haemoglobin decreased
		Hypoplastic anaemia
		Microcytic anaemia
		Proerythroblast count decreased
		Red blood cell count decreased
		Reticulocyte count decreased
		Reticulocytopenia
Effects on cardiac tissue	Cardiac failure effects	Acute left ventricular failure
		Acute pulmonary oedema
		Acute right ventricular failure
		Cardiac asthma
		Cardiac failure
		Cardiac failure acute
		Cardiac failure congestive
		Cardiac failure high output
		Cardiogenic shock
		Cardiopulmonary failure
		Cor pulmonale
		Cor pulmonale acute
		Ejection fraction decreased
		Left ventricular failure
		Low cardiac output syndrome
		Obstructive shock
		Pulmonary oedema
		Right ventricular failure
		Ventricular failure
Effects on skeletal muscle	Muscle related effects	Muscle necrosis
		Myoglobin blood increased
		Myoglobin blood present
		Myoglobin urine present
		Myoglobinaemia

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Myoglobinuria
		Myopathy
		Myopathy toxic
		Necrotising myositis
		Rhabdomyolysis
		Biopsy muscle abnormal
		Electromyogram abnormal
		Muscle disorder
		Muscle fatigue
		Muscular weakness
		Musculoskeletal discomfort
		Musculoskeletal disorder
		Musculoskeletal pain
		Myalgia
		Myositis
		Neck pain
		Dropped head syndrome
Infection events	Infections	All PTs in the 'Infections and Infestations' SOC
Infection events	Skin infection	Body tinea
		Skin candida
		Cellulitis
		Erysipelas
		Erysipeloid
		Folliculitis
		Fungal skin infection
		Furuncle
		Impetigo
		Herpes dermatitis
		Postoperative wound infection
		Skin infection
		Soft tissue infection

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Staphylococcal skin infection
		Staphylococcal infection
		Tinea infection
		Tinea capitis
		Tinea manuum
		Wound infection
		Wound infection bacterial
Infection events	Respiratory infection	Atypical pneumonia
		Bronchitis
		Empyema
		Influenza
		Laryngitis viral
		Lower respiratory tract infection
		Lower respiratory tract infection bacterial
		Lung abscess
		Lung infection
		Nasopharyngitis
		Pleural infection
		Pneumonia necrotising
		Post procedural pneumonia
		Pyopneumothorax
		Rhinitis
		Sputum purulent
		Pharyngitis
		Pneumocystis jirovecii infection
		Pneumonia
		Pneumonia bacterial
		Pneumonia fungal
		Pneumonia haemophilus
		Pneumonia pseudomonal
		Pneumonia staphylococcal

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Pneumonia viral
		Pulmonary sepsis
		Respiratory tract infection
		Sinusitis
		Tonsilitis
		Tracheitis
		Upper respiratory tract infection
Effects on the liver and related tissues	Transaminase and bilirubin elevations	Alanine aminotransferase increased
		Alanine aminotransferase abnormal
		Aspartate aminotransferase increased
		Aspartate aminotransferase abnormal
		Bilirubin conjugated increased
		Blood bilirubin increased
		Blood bilirubin unconjugated increased
		Hyperbilirubinaemia
		Hypertransaminasaemia
		Transaminases increased
Investigations	Investigations	Blood creatine phosphokinase increased
		Blood creatine phosphokinase abnormal
Effects on the gastrointestinal tract	Nausea	Nausea
Effects on the gastrointestinal tract	Vomiting	Vomiting
Effects on the gastrointestinal tract	Diarrhoeas	Diarrhoea
		Frequent bowel movements
General disorders	Asthenia effects	Asthenia
		Fatigue
General disorders	Peripheral oedemas	Lymphoedema
		Oedema
		Oedema peripheral
General disorders	Facial oedemas	Eyelid oedema
		Face oedema

Table 11 Adverse Events of Special Interest and grouped terms

Category	AESI Concept	MedDRA PT
		Periorbital oedema

AESI adverse event of special interest, CSR Central serous retinopathy, MedDRA Medical Dictionary for Regulatory Activities, PT preferred term, RPED Retinal Pigment Epithelial Detachment, SOC system organ class

Sort order for AESIs outputs:

All groupings below should appear in the AESI outputs in the following order

- Effects on the skin and mucous membranes:
  - Rash acneiform
  - Rash, non-acneiform
  - Rashes
  - Dry skin effects
  - Paronychia effects
  - Oral mucositis effects
- Effects on the eyes:
  - Retinal vein occlusion effects
  - Retinal pigmented epithelium detachment (RPED) / Central serous retinopathy (CSR) effects
  - Other retinal effects
- Effects on the lungs:
  - Interstitial lung disease (ILD-type) effects
  - Dyspnoea effects
- Effects on blood cells:
  - Febrile neutropenia effects
  - Neutropenia effects

  - Thrombocytopenia effects
  - Erythropenia effects
- Effects on cardiac tissue:
  - Cardiac failure effects
- Effects on skeletal muscle:
  - Muscle related effects
- Infection events:
  - Infections
  - Skin infection
  - Respiratory infection
- Effects on liver and related tissues:
  - Transaminase and bilirubin elevations
- Investigations:
  - Investigations
- Effects on the gastrointestinal tract:
  - Nausea
  - Vomiting
  - Diarrhoeas
- General disorders:
  - Asthenia effects
  - Peripheral oedemas
  - Facial oedemas