

Official Title: A Randomised, Parallel, Double Blinded Study to Compare the Efficacy and Safety of FKB238 to Avastin® In 1st Line Treatment for Patients with Advanced/Recurrent Non-Squamous Non-Small Cell Lung Cancer in Combination of Paclitaxel and Carboplatin AVANA

NCT Number: NCT02810457

Document Date: SAP Version 2.0: 07 May 2019

PAREXEL International

Centus Biotherapeutics

FKB238-002

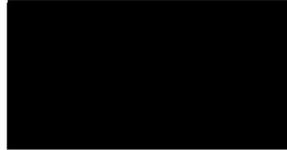
A Randomised, Parallel, Double Blinded Study to Compare the Efficacy and Safety of
FKB238 to Avastin® In 1st Line Treatment for Patients with Advanced/Recurrent
Non-Squamous Non-Small Cell Lung Cancer in Combination of Paclitaxel and
Carboplatin (AVANA)

Statistical Analysis Plan

PAREXEL Project Number: 227454

SPONSOR SIGNATURE PAGE

Approved by:



Head of Statistics
Centus Biotherapeutics

7 March 2019

Date


Associate Director in Statistics
MedImmune, LLC

Date


Principle Clinical Pharmacokineticist
MedImmune, LLC

Date

SPONSOR SIGNATURE PAGE

Approved by:


Head of Statistics
Centus Biotherapeutics

Date


Associate Director in Statistics
MedImmune, LLC

07 Mar 2019
Date


Principle Clinical Pharmacokineticist
MedImmune, LLC

07 May 2019
Date

PAREXEL SIGNATURE PAGE

This document has been approved and signed electronically on the final page by the following:

Signatory	
Author	 Project Role: Biostatistics Lead

TABLE OF CONTENTS

TABLE OF CONTENTS	4
LIST OF TABLES.....	5
LIST OF ABBREVIATIONS	6
1 INTRODUCTION.....	8
2 STUDY OBJECTIVES.....	8
3 INVESTIGATIONAL PLAN	8
3.1 Overall Study Design and Plan.....	8
3.2 Efficacy and Safety Variables	10
4 STATISTICAL METHODS	11
4.1 Data Quality Assurance.....	11
4.2 General Presentation Considerations.....	11
4.2.1 Definition of Baseline.....	12
4.3 Visit Windows	13
4.4 Study Patients	14
4.4.1 Disposition of Patients.....	14
4.4.2 Protocol Deviations	14
4.5 Analysis Populations	16
4.6 Demographic Data and Other Baseline Characteristics	16
4.7 Medical History	18
4.8 Prior and Concomitant Medications.....	18
4.9 Prior, Concomitant and Post-Treatment Anti-Cancer Therapies	19
4.10 Efficacy Evaluation	20
4.10.1 Analysis and Data Conventions.....	21
4.10.1.1 Multi-centre Studies.....	21
4.10.1.2 Adjustments for Covariates.....	21
4.10.1.3 Handling of Dropouts or Missing Data.....	22
4.10.1.4 Multiple Comparisons/Multiplicity.....	22
4.10.1.5 Interim Analyses	22
4.10.1.6 Examination of Subgroups.....	22
4.10.1.7 Investigator-Based Overall Response Assessments.....	22
4.10.2 Primary Efficacy Variable – Overall Response Rate	26
4.10.3 Secondary Efficacy Variables.....	29
4.10.3.1 Overall Response Rate at Week 19.....	29
4.10.3.2 Progression-free Survival.....	29
4.10.3.3 Overall Survival	33
4.10.3.4 Duration of Response.....	34
4.10.3.5 Disease Control Rate.....	35
4.10.4 Tumour and Response Assessments.....	35
4.10.5 Efficacy Endpoints Summary	35
4.11 Safety Evaluation.....	36
4.11.1 Extent of Exposure	36
4.11.1.1 Exposure and Dose Intensity.....	36
4.11.1.2 Delay and Dose Modification	39
4.11.2 Adverse Events	39
4.11.3 Adverse Events of Special Interest.....	42
4.11.4 Deaths and Other Significant Events.....	42
4.11.5 Clinical Laboratory Evaluation.....	43

4.11.6	Vital Signs, Physical Findings and Other Observations Related to Safety	44
4.11.7	Safety Monitoring (Independent Data Monitoring Committee)	45
4.12	Other Analyses	45
4.12.1	Immunogenicity	46
4.12.2	Pharmacokinetics	46
4.13	Determination of Sample Size	47
4.14	Changes in the Conduct of the Study or Planned Analysis	48
5	REFERENCES	48

LIST OF TABLES

Table 1	Overall Visit Response	23
Table 2	Evaluation of Target Lesions	23
Table 3	Schedule of Pharmacokinetic and Antidrug Antibody Sampling	46

LIST OF ABBREVIATIONS

ADA	antidrug antibody
AE	adverse event
ALK	anaplastic lymphoma receptor tyrosine kinase
ALT	alanine transaminase
AST	aspartate transaminase
ATC	anatomical therapeutic chemical
AUC	area under curve
BICR	blinded independent central review
BL	baseline
BLQ	below lower limit of quantification
BMI	body mass index
BOR	best overall response
BSA	body surface area
CI	confidence interval
C _{max}	serum maximum concentration
CR	complete response
CRO	clinical research organization
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
C _{trough}	serum trough concentration
CV	coefficient of variation
DCR	disease control rate
DOR	duration of response
ECG	electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	electronic case report form
EGFR	epidermal growth factor receptor
EMA	European Medicines Agency
EU	European Union
FDA	Food and Drug Administration
GFR	glomerular filtration rate
hCG	human chorionic gonadotropin
IDMC	independent data monitoring committee
IP	investigational product; FKB238 and EU-Avastin
ITT	intent-to-treat
LLOQ	lower limit of quantification
LVEF	left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
NA	not applicable
nAb	neutralising antibodies

NC	not calculable
NCI	National Cancer Institute
NE	not evaluable
NED	no evidence of disease
NS-NSCLC	non-squamous non-small cell lung cancer
NTL	non-target lesion
ORR	overall response rate
OS	overall survival
PD	disease progression (progression of disease, progressive disease)
PDS	protocol deviation specification
PFS	progression-free survival
PK	pharmacokinetics
PPS	per protocol set
PR	partial response
PT	preferred term
RECIST	Response Evaluation Criteria in Solid Tumours
RTSM	Randomisation and Trial Supply Management
SAE	serious adverse event
SAP	statistical analysis plan
SCLC	small cell lung cancer
SD	stable disease
SI	International System of Units
SOC	system organ class
TEAE	treatment-emergent adverse event
TL	target lesion
TOST	two one-sided tests
ULN	upper limit of normal
WHO-DD	World Health Organization Drug Dictionary

1 INTRODUCTION

The aim of the Statistical Analysis Plan (SAP) on hand is to state the scope of analyses and to describe the planned analyses for protocol FKB238-002, a randomised, parallel, double blinded study to compare the efficacy and safety of FKB238 to EU-Avastin® in 1st line treatment for patients with advanced/recurrent non-squamous non-small cell lung cancer (NS-NSCLC) in combination of paclitaxel and carboplatin.

The current version 2 of the SAP is based on version 5 of the study protocol, dated 22-May-2018.

In addition, a post-text table, listing and figure shells document will be created to complement this SAP.

2 STUDY OBJECTIVES

The primary objective of this study is:

- To demonstrate the efficacy equivalence of FKB238 and EU-Avastin when used in combination with paclitaxel/carboplatin as measured by overall response rate (ORR)

Secondary objectives of this study are:

- To compare FKB238 and EU-Avastin through:
 - ORR at Week 19
 - Progression-free survival (PFS)
 - Overall survival (OS)
 - Duration of response (DOR)
 - Disease control rate (DCR)
- To compare the safety of FKB238 and EU-Avastin
- To compare the antidrug antibodies (ADAs) produced by FKB238 and EU-Avastin
- To compare the serum trough concentration (C_{trough}) of FKB238 and EU-Avastin

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This is a global multi-centre, double-blind, parallel comparative, Phase 3 study designed to compare the efficacy and safety of FKB238 and EU-Avastin when used in combination with paclitaxel and carboplatin in the 1st line treatment of advanced or recurrent NS-NSCLC. The primary objective of this study is to demonstrate the efficacy equivalence of FKB238 and EU-Avastin when used in combination with paclitaxel/carboplatin as measured by ORR.

Patients aged 18 years or older who have advanced or recurrent NS-NSCLC will be screened for participation up to 28 days before randomisation. Approximately 730 eligible patients will be randomly assigned in a 1:1 ratio to 1 of 2 treatment groups. Randomisation will be stratified according to epidermal growth factor receptor (EGFR) mutation and anaplastic lymphoma receptor tyrosine kinase (ALK) gene arrangement status (both are tested and known negative versus status unknown for either), geographical region (North America, Western Europe, East Asia, All Other Regions),

prior weight loss over the previous 6 months (<5% yes versus no) and disease stage (advanced or recurrent).

The following treatment groups will be included in this study:

- FKB238 group: paclitaxel + carboplatin (combination drugs) + FKB238 (IP)
- Avastin group: paclitaxel + carboplatin (combination drugs) + EU-Avastin (IP)

Upon randomisation, patients will enter the study treatment period. Following start of study treatment, patients will attend study visits every 3 weeks as long as they are receiving study treatment. The combination drugs (paclitaxel + carboplatin) will be administered on Day 1 of each 21-day cycle for at least 4, and no more than 6 cycles. The number of cycles is determined by patient's need and the investigator's assessment. FKB238 or EU-Avastin investigational products (IP) will also be administered on Day 1 of each 21-day cycle until objective progressive disease (PD) or other criteria for treatment discontinuation are met.

Radiological assessment using Response Evaluation Criteria in Solid Tumours (RECIST) v1.1 will be performed at screening and every 6 weeks (± 1 week) for 24 weeks, then every 9 weeks (± 1 week) for the remainder of the treatment period. Patients who discontinue study treatment for reasons other than PD will continue to undergo tumour assessments every 12 weeks (± 1 week) until PD, death or the data cut-off (whichever occurs first).

Patients who develop toxicity may delay study treatment for up to 3 weeks. Patients who require the combination drugs and/or IP to be held for more than 3 weeks will discontinue the study treatment.

After the discontinuation of the study treatment, further care and treatment will be at the discretion of the investigator. Any systemic anti-cancer treatment, radiotherapy or anti-cancer surgery conducted after discontinuation of study treatment will be collected until death, loss to follow-up, withdrawal of consent or until end of study. Assessments for survival should be made every 8 weeks (± 1 week) following objective PD.

Data cut-off is defined as 12 months from randomisation of the last patient enrolled for the purpose of the analyses.

The Schedule of Assessments indicates the timing of the planned visits including protocol defined visit window for the study visits (see Schedule of Assessments table from the protocol). The investigator/sub-investigator should adhere to the Schedule of Assessments procedures and perform tests/observations according to the protocol. Assessments after start of study treatment should be performed within a window of ± 3 days of the scheduled visit date. Computed tomography (CT) or magnetic resonance imaging (MRI) assessments should be done within a window of ± 7 days of the scheduled date according to the RECIST v1.1 criteria, but every effort should be made to adhere as closely as possible to the original schedule of scans.

Investigators, site staff including pharmacy staff, patients, clinical research organization (CRO) personnel, and Sponsor personnel (except for a specified IP distribution manager) will be blinded to individual patient treatment assignment during the course of the study.

ClinPhone®RTSM (Randomisation and Trial Supply Management) will develop the randomisation schedule.

The IP will be packaged and labelled in such a way that visual inspection of the IP or packaging will not reveal the treatment assignment; however, each individual kit of IP will be numbered so that, if necessary, the number can be used to break the treatment blind if this becomes necessary to protect the safety of the patient.

3.2 Efficacy and Safety Variables

RECIST v1.1 criteria will be used to assess patient response to treatment by determining ORR, DOR, PFS, and DCR.

Radiological imaging will be assessed by investigators as well as an independent central radiological committee based on RECIST v1.1 criteria. Primary efficacy assessment will be presented based on the investigators' assessment results as well as the independent central radiological assessment results, but primary efficacy assessment based on the independent central radiological assessment results will be used as a pivotal result of the study.

Baseline RECIST v1.1 assessment will be performed using CT (or MRI scans where CT is clinically contra-indicated) at screening no more than 28 days prior to start of study treatment (Cycle 1 Day 1) and ideally should be performed as close as possible to the start of the study treatment. Baseline radiological assessments should cover the chest and the upper abdomen (including liver and the adrenal glands). Any other areas of the disease involvement should be additionally investigated based on signs and symptoms of individual patients. Brain CT/MRI is required within 4 weeks prior to the randomisation.

After start of study treatment, RECIST v1.1 assessment using CT (or MRI scans where CT is clinically contra-indicated) will be performed every 6 weeks (± 1 week) for 24 weeks, then every 9 weeks (± 1 week) for the remainder of the treatment period, until objective PD. The methods of assessment of tumour burden used at baseline CT or MRI scans of chest and upper abdomen (including liver and adrenal glands) must be used at each subsequent follow-up assessment.

No further radiological tumour assessments will be done after objective PD is documented. Patients who discontinue study treatment for reasons other than PD will continue to undergo RECIST v1.1 tumour assessments every 12 weeks (± 1 week) until PD, death or the data cut-off (whichever occurs first).

Adverse events (AEs) will be collected from time of signature of informed consent, throughout the treatment period and up to and including 30 calendar days after the last study treatment. All ongoing and any new AEs/Serious AEs (SAEs) identified during the 30 calendar days after last dose of study treatment must be followed to resolution unless the event is considered by the investigator to be unlikely to resolve, or the patient is lost to follow-up.

All pregnancies and outcomes of pregnancy should be reported during the course of the study and within 30 days of the last dose of study treatment. A serum/urine pregnancy human chorionic gonadotropin (hCG) test will be performed for women of childbearing

potential at screening. If a patient becomes pregnant during the course of the study, study treatment should be discontinued immediately.

Blood and urine samples for determination of clinical chemistry, haematology and urinalysis will be taken pre-dose at the times indicated in the Schedule of Assessments table from the protocol. If clinical chemistry, haematology and urinalysis screening assessments have been performed within 14 days prior to starting study treatment, they do not need to be repeated on Cycle 1 Day 1 if the patient's condition has not changed (no new treatment during this period of time, no new complication or worsening).

Additional safety laboratory samples may be collected if clinically indicated at the discretion of the investigator. The date and results (values, units and reference ranges) will be recorded on the appropriate electronic case report form (eCRF).

Vital signs will be assessed according to the Schedule of Assessments table from the protocol and will include systolic and diastolic blood pressure (mmHg), pulse rate (beats per minute), and body temperature (°C). Assessments will be performed pre-dose, at the visits as shown in the Schedule of Assessments table from the protocol, and at occurrence of any cardiac AE. Weight will be recorded at screening and then Day 1 of each cycle as well as the Discontinuation Visit. Height will be assessed at screening only. Physical examinations will be conducted according to the Schedule of Assessments table from the protocol. Electrocardiograms (ECGs) are required within 7 days prior to starting study treatment, on Day 1 of Cycle 1, 2 hours within pre-dose and post-dose, when clinically indicated and at the Follow-up Visit after the patient has discontinued study treatment. Patients should have an Eastern Cooperative Oncology Group (ECOG) Performance Status (PS) of 0 or 1 at Screening to be included in the study. The investigator will assess ECOG PS at Screening, and during the Study Treatment Period, and Follow-up Period as noted in the Schedule of Assessments table from the protocol.

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard PAREXEL procedures.

4.2 General Presentation Considerations

Later sections of the SAP use the treatment discontinuation date. This is the date of withdrawal from study treatment. In general it is not the same as the date of the last dose of study treatment.

Continuous data will be summarised in terms of the mean, standard deviation, median, minimum, maximum and number of observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The standard deviation will be reported to two more decimal places than the raw data recorded

in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarised in terms of the number of patients providing data at the relevant time point (n), frequency counts and percentages. Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Unless otherwise stated, the denominator for percentage calculations will be the number of patients in the analysis population. The missing category will only be presented in the tables if events / patients are observed in this category.

Changes from baseline in categorical data will be summarised using shift tables where appropriate.

Safety assessment values of the form of " $<x$ " (i.e., below the lower limit of quantification) or " $>x$ " (i.e., above the upper limit of quantification) will be imputed as " x " in the calculation of summary statistics but displayed as " $<x$ " or " $>x$ " in the listings. Values of " $\leq x$ " or " $\geq x$ " will be imputed as well.

Summaries based on the Intent-to-treat (ITT) population will summarise data as overall and by randomised treatment (except for efficacy summaries where data will be summarised by treatment arm only). Summaries based on the Safety population, the Per Protocol Set (PPS) and the PK population will summarise data as overall and by actual treatment.

All report outputs will be produced using SAS version 9.3 or higher.

4.2.1 Definition of Baseline

In general, for efficacy endpoints the last observed measurement prior to randomisation will be considered the baseline measurement.

For safety analyses, baseline is defined as the last non-missing assessment made before the first dose of study treatment (IP and/or combination drugs) in the study.

Safety assessments done on the day of first dose of study treatment (Day 1) will be handled as follows:

- ECG assessments are planned pre-dose and post-dose on Day 1. The "pre-dose"/"post-dose" indicator will be used to identify assessments prior to first dose. In case of a missing "pre-dose"/"post-dose" indicator the assessment time will be compared with the time of the first dose to identify pre-dose assessments.
- For non-ECG safety assessments both dates and times will be used to check if an assessment was done prior to the first dose. If either the assessment time or the time of the first dose is not available, then it will be assumed that assessments on Day 1 are pre-dose, if such assessments are required by the protocol to be conducted before first dose. This is required per protocol for clinical chemistry, haematology, urinalysis, coagulation, vital signs, physical examination and the ECOG performance status.

For vital signs, the average of the three blood pressure screening values will be used as baseline, if no later values prior to start of study treatment are available.

4.3 Visit Windows

For safety data to be summarised by visit (i.e. safety lab and vital signs), all assessments will be assigned to calculated visit windows (using study day).

For the purpose of safety data summary, study day 1 will be defined as the date of first dose of study treatment. For assessments that occur on or after first dose, study day will be defined as (date of assessment - date of first dose of study treatment + 1). For assessments that occur prior to first dose, study day will be defined as (date of assessment - date of first dose of study treatment). There is no study day 0.

The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the visit window should be based on the actual date of the assessment. For summaries at a subject level, all values should be included, regardless of whether they appear in a corresponding visit based summary, when deriving a subject level statistic such as a maximum.

The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.-

The following time windows will be used for safety lab variables and the vital sign variables blood pressure, pulse and temperature to remap the visits:

Visit	Target Day	Visit Window
Cycle 1 Day 8	Day 8	Day 2 to Day 11
Cycle 1 Day 15	Day 15	Day 12 to Day 18
Cycle 2 Day 1	Day 22	Day 19 to Day 32
Cycle 3 Day 1	Day 43	Day 33 to Day 53
Cycle 4 Day 1	Day 64	Day 54 to Day 74
...
Cycle n Day 1 (n≥3)	Day (n-1)*21+1	Target Day ± 10 days

For weight and BMI no assessments at Cycle 1 Day 8 and Cycle 1 Day 15 are planned, and the following visit windows will be used in the remapping:

Visit	Target Day	Visit Window
Cycle 2 Day 1	Day 22	Day 2 to Day 32
Cycle 3 Day 1	Day 43	Day 33 to Day 53
Cycle 4 Day 1	Day 64	Day 54 to Day 74
...
Cycle n Day 1 (n≥3)	Day (n-1)*21+1	Target Day ± 10 days

For visit-based summaries, if there is more than one value per subject within a visit window then the closest to the planned study day value should be summarised, or the earlier in the event the values are equidistant from the planned study day.

Listings should display all values contributing to a time point for a subject; they should also highlight the value for that subject that was used in the summary table, wherever feasible.

4.4 Study Patients

4.4.1 Disposition of Patients

A clear accounting of the disposition of all patients who enter the study will be provided, from screening to study completion.

Study Populations

A summary of the number and percentage of patients from the following categories will be tabulated:

- ITT population
- PPS
- Safety population
- PK population
- ADA evaluable population

ITT Population

A summary of the number of patients randomised per centre, per country and per region will be tabulated using the ITT population.

Patient Disposition

Patient disposition will be summarised by randomised treatment group (when applicable) and overall.

Summary will include the number and percentage of patients:

- Screened
- Screen failures
- Randomised
- Randomised but not treated (IP or combination drugs)
- Randomised and treated
- Discontinued from treatment and reasons for treatment discontinuation
- Still in the study at data cut-off
- Discontinued from the study and reasons for study discontinuation

In addition, following listings will be provided:

- A by-patient listing of patient disposition data based on all screened patients
- A by-patient listing of end of study treatment based on the ITT population
- A by-patient listing of end of study data based on the ITT population

4.4.2 Protocol Deviations

The important protocol deviations will lead to exclusion from the PPS. This section describes important deviations related to the study inclusion or exclusion criteria, trial conduct, patient management or patient assessments, which have a potential impact on the efficacy outcomes.

The following list of important protocol deviations consists of 6 selection criteria not adhered to plus 5 specific deviations:

The Selection Criteria to be Adhered to:

- Newly diagnosed advanced (stage IV) /recurrent NS-NSCLC for which they had not received any systemic anti-cancer therapy for metastatic disease, including

- chemotherapy, biologic therapy, immunotherapy, or any investigational drug (Inclusion Criteria No. 2).
- Histologically or cytologically confirmed diagnosis of predominantly NS-NSCLC (Inclusion Criteria No. 3).
 - Be eligible to receive study treatment of bevacizumab, paclitaxel, and carboplatin for the treatment of advanced or recurrent NS-NSCLC (Inclusion Criteria No. 4).
 - Existence of at least 1 measurable lesion by response evaluation criteria (RECIST v1.1), defined as; at least one lesion, not previously irradiated, that can be accurately measured at baseline as ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with CT or MRI which is suitable for accurate repeated measurements (Inclusion Criteria No. 5).
 - Small cell lung cancer (SCLC) or combination SCLC and NSCLC. Squamous-cell tumours and mixed adenosquamous carcinomas of predominantly squamous nature (Exclusion Criteria No. 1).
 - Recurrence occurred within 12 months from the last dose of neoadjuvant/adjuvant therapy (Exclusion Criteria No. 2).

Specific Deviations:

- Baseline imaging >28 days prior to randomisation.
- Subjects randomised but who did not receive any IP.
- Misrandomisations and errors in treatment dispensing, includes 2 categories:
 - Subjects who delayed study treatment for more than 21 days due to any reason.
 - The subject received the incorrect randomised therapy intermittently, i.e. actually received both FKB238 and Avastin at least once.

If a subject actually received the wrong IP throughout the study, i.e. randomised to FKB238 but always received Avastin or vice versa, then this is not considered as an important protocol deviation and the patient is still included in the PPS in the actual treatment group, provided that he/she has no other important protocol deviations. If a patient got wrong kits, but all kits still included the correct randomised IP, then this is also not considered as an important protocol deviation.

- No baseline RECIST v1.1 assessment on or before date of randomisation
- Anti-cancer therapy received prior to study treatment discontinuation
-

If any patients have any other protocol deviations which are considered important by the study team, these latter will also lead to exclusion from the PPS.

In addition to the programmatic determination of the deviations, monitoring notes or summaries will be reviewed to determine any important protocol deviations that are not identifiable via programming. The final classification will be made prior to database lock.

A summary table and listing of protocol deviations will be provided as follows:

- The number and percentage of patients without any important deviation, with at least 1 important protocol deviation, and type of important protocol deviations will be summarised by randomised treatment group and overall. (Analysis population: ITT population)
- Two by-patient listings of important protocol deviations and all protocol deviations will be presented. (Analysis population: ITT population)

4.5 Analysis Populations

The following analysis populations will be included for this study:

ITT Population: all patients randomised to treatment. Patients will be included in the treatment group according to the randomisation assigned, regardless of the treatment actually given. All efficacy analyses will be performed on the ITT population. These analyses will be treated as the primary analysis for the Food and Drug Administration (FDA) requirement and as sensitivity analysis otherwise.

Per Protocol Set: all patients randomised to treatment who received at least 1 dose of IP with no important protocol deviations (see Section 4.4.2). Patients will be included in the treatment group according to the treatment actually given. All efficacy analyses will be performed on the PPS. These analyses will be treated as the primary analysis for the European Medicines Agency (EMA) requirement and as sensitivity analysis otherwise.

The subjects to be included in the PPS will be identified in a Data Review Meeting which will be held prior to database lock for final analysis. The PPS will be approved by the Sponsor.

Safety Population: all patients randomised to treatment who received at least 1 dose of IP. Patients will be included in the treatment group according to the treatment actually given. All safety analyses will be performed on the Safety Population. Treatment groups will be analysed according to the IP actually received in the majority of kits.

PK Population: all PPS patients who have at least one serum drug concentration data, which is defined in the study protocol, after IP administration. Patients will be included in the treatment group according to the treatment actually given.

ADA Evaluable Population: all PPS patients who have at least one ADA assessment prior to and after baseline data. Patients will be included in the treatment group according to the treatment actually given.

A by-patient listing of analysis population details will be provided. This listing will include: centre, patient identifier and inclusion/exclusion flag for each population. This listing will be based on all patients screened.

4.6 Demographic Data and Other Baseline Characteristics

All summaries and listings described in this section are based on the ITT population.

Demographic data will be summarised using descriptive statistics and frequency tables, respectively:

- Age (years) continuous and categorical (<65, ≥65, additional subgroups 18-39, 40-64, 65-74, 75-84, ≥85), gender, ethnicity, race, weight (kg), height (cm) and body mass index (BMI) (kg/m²)

The BMI will be calculated as quotient of weight in kilograms divided by height in metres squared.

Further baseline data will be summarised using descriptive statistics and frequency tables, respectively:

- Pathology at diagnosis: histology type, initial stage and time from earliest diagnosis of lung cancer to randomisation (continuous)
- Extent of disease: disease stage (advanced, recurrent), overall disease classification (metastatic, locally advanced), sites of metastatic disease, sites of locally advanced disease
- Smoking history including number of pack years
- ECOG PS

The time from diagnosis of lung cancer to randomisation will be calculated in units of months as the difference between the date of randomisation and date of diagnosis of lung cancer plus 1, divided by 365.25 and multiplied by 12. Partial diagnosis dates will be imputed, inserting 15 for a missing day and inserting 01-Jul for missing month and day. After imputation, compare the imputed date with the randomisation date, and if the imputed date is >randomisation date, then replace it with the randomisation date.

Randomisation of patients in the study will be stratified according to the following stratification factors:

- EGFR mutation and ALK gene arrangement status: both are tested and known negative versus status unknown for either
- Geographical region: North America versus Western Europe versus East Asia versus All other regions
- Prior weight loss over the previous 6 months <5%: Yes versus No
- Disease stage: Advanced versus Recurrent

After randomisation, updates to the original stratification factors used for randomisation, for correction purposes, are possible. Consequently, both original and final stratification factors (post correction, if any) will be available and both will be displayed in a frequency table.

Summary of EGFR mutation and ALK gene arrangement status will also be displayed in a frequency table including both baseline status as well any post randomisation updates.

Baseline assessments of target lesions (TLs) and non-target lesions (NTLs) will be summarised in a summary table as follows:

- TLs: number and percentage of patients with at least one TL, number of TLs by patient, sum of diameters (mm), number of organ sites by patient, organ sites (number and percentage of patients with at least one TL at a site, by site)
- NTLs: number and percentage of patients with at least one NTL, number of NTLs per patient, number of organ sites per patient, organ sites (number and percentage of patients with at least one NTL at a site, by site)

By-patient listings of demographic data, other baseline characteristics (as summarised above) will be provided. Notably, a by-patient listing of height, weight and BMI assessments will also be provided. This listing will include both baseline and post-baseline assessments.

4.7 Medical History

Medical history conditions will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) and will be summarised by system organ class (SOC) and preferred term (PT). Missing coding terms will be listed and summarised as "Not coded".

Medical history conditions will be categorised as previous or current as follows:

- Previous: Medical history conditions which are specified as start date prior to the date of randomisation with no ongoing status or the stop date prior to the date of randomisation.
- Current: Medical history conditions which are specified as start date prior to the date of randomisation with ongoing status or stop date on or after the date of randomisation.

If medical history conditions start and/or stop dates are missing or partial, the dates will be compared as far as possible with the date of randomisation. Medical history conditions will be assumed to be current, unless there is clear evidence (through comparison of partial dates) to suggest that the medical history condition stopped prior to the date of randomisation.

The following summaries will be produced based on the ITT population:

- Summary of previous medical history conditions
- Summary of current medical history conditions

A by-patient listing of medical history data will be provided based on the ITT population.

4.8 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the most recent version of the WHO-DD and will be summarised by Anatomical Therapeutic Chemical (ATC) Class Level 1 (Anatomical Main Group), Level 3 (Pharmacological Subgroup), and PT. Missing coding terms will be listed and summarised as "Not coded".

Prior and concomitant medications will be defined based on start and stop dates as follows:

- Prior medications are those with a stop date prior to the date of randomisation
- Concomitant medications are those with a start date prior to, on or after the date of randomisation with ongoing status or with a stop date on or after the date of randomisation

If medication start and/or stop dates are missing or partial, the dates will be compared as far as possible with the date of randomisation. Medications will be assumed to be concomitant, unless there is clear evidence (through comparison of partial dates) to suggest that the medication stopped prior to the date of randomisation.

Concomitant medications will be summarised for the ITT population.

In addition, by-patient listings of prior and concomitant medications will be provided based on the ITT population.

Anti-cancer medications will be reported and summarised separately, see the next section.

4.9 Prior, Concomitant and Post-Treatment Anti-Cancer Therapies

All summaries and listings described in this section are based on the ITT population.

Prior Anti-Cancer Therapies

Prior anti-cancer therapies will be summarised as follows:

- Number and percentage of patients by class of anti-cancer therapies and by treatment status

Prior radiotherapy will be summarised as follows:

- Number and percentage of patients with any history of prior radiotherapy, by treatment status

Prior surgeries will be coded using the most recent version of the MedDRA. A summary of prior surgeries by SOC and PT will be provided. Missing coding terms will be listed and summarised as "Not coded".

By-patient listings of prior anti-cancer therapies, prior radiotherapy and prior surgeries will be provided.

Concomitant and Post-Treatment Anti-Cancer Therapies

No additional systemic anti-cancer treatment including other chemotherapies, radiotherapy, anti-cancer agents, and investigational agents is allowed to be used prior to discontinuation of study treatment. Palliative radiotherapy for painful bone metastases (except for thoracic region) are allowed during the study treatment. Herbal medications intended for anti-cancer use are prohibited.

Following discontinuation of study treatment, patients may receive any subsequent therapy for NS-NSCLC at the discretion of the investigator. After discontinuation of study treatment, all systemic anti-cancer treatment, radiotherapy or anti-cancer surgery will be collected until death, loss to follow-up, withdrawal of consent or until the data cut-off (whichever occurs first).

Summaries described hereafter in this section will not include palliative radiotherapy and will also not include surgeries which are not anti-cancer surgeries. These latter will only be listed. These categories of therapies will not be taken into account when defining post-treatment (also called subsequent) anti-cancer therapies. Only anti-cancer therapies, non-palliative radiotherapy and anti-cancer surgeries will be considered.

Concomitant and post-treatment anti-cancer therapy, radiotherapy or anti-cancer surgery will be defined based on start and stop dates as follows:

- Concomitant anti-cancer therapy, radiotherapy or anti-cancer surgery are those with a stop date on or after the first dose date of study treatment (IP or combination drugs) or with ongoing status (only applies to anti-cancer therapy) and with a start date prior to the study treatment discontinuation date.
- Post-Treatment anti-cancer therapy, radiotherapy or anti-cancer surgery are those with a start date on or after the study treatment (IP or combination) discontinuation date.

In patients randomised who did not start study treatment (IP or combination drugs), any anti-cancer therapy, radiotherapy or anti-cancer surgery with a stop date after the date of randomisation or with ongoing status (only applies to anti-cancer therapy) will be considered as post-treatment.

All post-treatment anti-cancer therapy, radiotherapy or anti-cancer surgery will be identified as “prior disease progression”, “post disease progression” or “no disease progression”, as follows:

- Any anti-cancer therapy, radiotherapy or anti-cancer surgery with a start date prior to the date of disease progression (see Section 4.10.3.2) will be considered as “prior disease progression”.
- Any anti-cancer therapy, radiotherapy or anti-cancer surgery with a start date after or on the date of disease progression will be considered as “post disease progression”.
- Any anti-cancer therapy, radiotherapy or anti-cancer surgery given to a subject who does not experience disease progression will be considered as “no disease progression”.

Of note the date of disease progression considered will be the one based on the independent central radiological assessments.

A summary of post-treatment anti-cancer therapies will be provided presenting number and percentage of patients by type of treatment (cancer therapies, radiotherapy and surgery) and by period (prior disease progression, post disease progression, no disease progression). This summary will also present time from discontinuation of study treatment to initiation of post-treatment anti-cancer therapy. This time will be calculated in units of days as the difference of the two dates (date of initiation of post-treatment anti-cancer therapy – stop date of study treatment + 1). The earliest start date of any post-treatment anti-cancer therapies will be considered, irrespective of the type of treatment.

Cancer surgeries will be coded using the most recent version of the MedDRA. A summary of post-treatment anti-cancer surgeries by category (prior disease progression, post disease progression, no disease progression, overall) and by SOC and PT will be provided. Missing coding terms will be listed and summarised as "Not coded".

By-patient listings of concomitant and post-treatment anti-cancer therapies, concomitant and post-treatment radiotherapy and concomitant and post-treatment surgeries will be provided.

4.10 Efficacy Evaluation

The primary and secondary endpoints of this study will be evaluated and presented based on the ITT population and the PPS.

Radiological imaging will be assessed by investigators as well as an independent central radiological committee based on RECIST v1.1 criteria. Primary efficacy assessment will be presented based on the investigators’ assessment results as well as the independent central radiological assessment results, but primary efficacy assessment based on the independent central radiological assessment results will be used as a pivotal result of the study.

Of note, programmatic derivation of RECIST v1.1 TLs response and overall response of subjects at visits will be performed as detailed in Section 4.10.1.7 based on the tumour assessments data entered into the eCRF. These data will be used when presenting efficacy endpoints based on the investigators' assessments. Overall response assessment as entered into the eCRF by investigators will only be listed.

Unless otherwise specified, study day 1 will be defined as the date of randomisation for efficacy analysis.

4.10.1 Analysis and Data Conventions

This study is designed to test for equivalence. The Null hypothesis for the treatment comparison will be that there is non-equivalence between EU-Avastin and FKB238 in ORR. The alternative hypothesis will be that there is equivalence. Different testing procedures will be applied to meet EMA's requirement and FDA's requirement.

EMA Approach

A two one-sided test procedure will be used to test this hypothesis. The analysis will be performed using the PPS. A 95% CI for the ORR difference between FKB238 and EU-Avastin will be estimated and compared to the margin [± 0.1221], which is deemed to represent a clinically acceptable difference with respect to ORR. If the true CI is within the interval [± 0.1221], an equivalence between FKB238 and EU-Avastin, with respect to the ORR, is confirmed.

FDA Approach

A two one-sided test procedure will be used to test this hypothesis. The analysis will be performed using the ITT population. A 90% CI for the ORR ratio between FKB238 and EU-Avastin will be estimated and compared to the margin [0.73, 1.38], which is deemed to represent a clinically acceptable difference with respect to ORR. If the true CI is within the interval [0.73, 1.38], an equivalence between FKB238 and EU-Avastin, with respect to the ORR, is confirmed.

4.10.1.1 Multi-centre Studies

For the purpose of the summaries and analyses, the term 'Centre' will be used to define each site.

Centre effect will not be included in any statistical model.

Descriptive summaries will not be presented by individual centre. Country and centre information will be included in the patient data listings.

4.10.1.2 Adjustments for Covariates

The following covariates will be considered for adjustment in analysis of primary and secondary endpoints:

- EGFR mutation and ALK gene arrangement status
 - Both are tested and known negative
 - Unknown for either
- Geographical region
 - North America
 - Western Europe

- East Asia
 - All Other Regions
- Prior weight loss over the previous 6 months (<5%)
 - Yes
 - No
- Disease stage
 - Advanced
 - Recurrent
- ECOG PS at baseline
 - 0
 - 1
- Gender
 - Male
 - Female
- Smoking history
 - Never
 - Former
 - Current
- Age
 - <65 years
 - ≥65 years

The covariates based on the stratification factors will be defined using the original stratification factor values, i.e. the values used for randomisation.

4.10.1.3 Handling of Dropouts or Missing Data

Unless specified, no imputation of missing data will be performed.

4.10.1.4 Multiple Comparisons/Multiplicity

One primary variable has been defined for this study. The equivalence of EU-Avastin and FKB238 will be tested using two different procedures. However each approach will be used separately with EMA decision solely based on the EMA approach and FDA decision solely based on the FDA approach. The secondary variables defined are intended to provide supportive evidence relating to the primary objective. No interim analyses are planned. Hence no adjustments for multiplicity are required.

4.10.1.5 Interim Analyses

No interim analyses are planned.

4.10.1.6 Examination of Subgroups

No subgroup analyses are planned.

4.10.1.7 Investigator-Based Overall Response Assessments

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

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

The Table 1 below provides the derivations of overall visit response using the information from TL, NTL and new lesions.

Table 1 Overall Visit Response

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

CR=complete response, NA=not applicable, NE=not evaluable, NED=no evidence of disease, PD=progressive disease, PR=partial response, SD=stable disease. NED is only applicable for the blinded independent central review.

Response of NTLs and presence of new lesions will be directly recorded into the eCRF. Response of TLs will be programmatically derived from the TLs assessments data recorded into the eCRF.

Derivation of Target Lesions (TLs) Responses

The rules defined in the Table 2 below will be used to determine the TLs response at any given visit. Derivation of TLs visit response will not be performed at visits after an overall response of PD.

Table 2 Evaluation of Target Lesions

CR	Disappearance of all TLs since baseline. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10 mm.
PR	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters, and criteria for PD are not met.

SD	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as a reference the smallest sum of diameters of TLs while on study.
PD	At least a 20% increase in the sum of diameters of TLs, taking as reference the smallest sum of diameters on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum of diameters must also demonstrate an absolute increase of at least 5mm. (Note: the appearance of one or more new lesions is also considered progression).
NE	Only relevant if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit. Note: If the sum of diameters meets the PD criteria, PD overrides NE as a TL response.

CR=complete response, NE=not evaluable, PD=progressive disease, PR=partial response, SD=stable disease, TL=target lesion.

The sum of diameters will be calculated as the sum of recorded sizes of diameter. The nadir sum of diameters at a visit will be calculated as the minimum of the sum of diameters measured within the set of tumour assessments up until and including this visit.

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to 1 decimal place before assigning a TL response. For example, 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

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

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

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

If any TL measurement is missing then the TL visit response is Not Evaluable. However, if the sum of non-missing TL diameters would result in PD (i.e. if using a value of 0 for missing TLs the sum of diameters has still increased by 20% or more from nadir and has an absolute increase of ≥ 5 mm), PD takes precedence over NE. The overall visit response will also be NE, unless there is a progression of the non-missing TLs or the NTLs, or any new lesions, in which case the response will be PD.

Lymph Nodes

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

TL Visit Responses Subsequent to CR

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

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes), the response will be set to CR irrespective of whether, when referencing the sum of TL diameters, the criteria for PD is also met.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes), the response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD is also met.
- Step 3: If not all lesions meet the CR criteria, and the sum of diameters meets the criteria for PD, the response will be set to PD.
- Step 4: If, after steps 1 through 3, a response cannot be determined, the response will be set to remain as CR.

TL too Big to Accurately Measure

If a TL becomes too big to accurately measure, this should be indicated in the database and an estimated size ('x') above which it cannot be accurately measured should be recorded. In this case (including where 'x' is missing) a target lesion visit response of PD will be assigned. If using a value of x in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team blinded to treatment assignment.

TL too Small to Accurately Measure

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

Lesion Intervention

Any TL (including lymph nodes), which has had intervention during the study should be handled in the following way:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.
- Step 2: If there was no evidence of progression after step 1, then a scaled sum of diameters will be calculated, treating the lesion diameter (for those lesions with intervention) as missing and replacing it by the lesion diameter at the nadir visit as long as $\leq 1/3$ of the TL measurements are missing because of intervention. If $> 1/3$ of TL measurements are missing because of intervention then TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e. if using a value of 0 for missing lesions, the sum of diameters has still increased by 20% or more compared to nadir and the sum of TLs has increased by ≥ 5 mm from nadir). If the scaling results in a visit response of PD then the patient will be assigned a TL response of PD. This step can only be applied where there are no other TLs with missing measurement other than because of intervention. If there are any other missing TLs measurement and there is no evidence of progression after step 1, the TL visit response will be assigned NE.
- Step 3: If after both steps PD has not been assigned, then the scaled sum of diameters (as long as $\leq 1/3$ of the TL measurements are missing because of intervention) will be

used to assign a visit response of PR or SD. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention also have a value of 0 recorded.

At subsequent visits the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with intervention should be treated as missing and scaled up (as per step 2 above). Once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours.

TLs that Split in Two

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

TLs that Merge

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

Change in Method of Assessment of TLs

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

4.10.2 Primary Efficacy Variable – Overall Response Rate

A blinded independent central review (BICR) of all scans used in the RECIST v1.1 assessment of tumour response will be conducted. Up to three qualified radiologists will independently review the imaging scans in the following way. First a primary review will be performed by two independent radiologists for each patient, on a time point by time point basis to give an overall tumour assessment at each time point using RECIST 1.1. Then a global radiology review will be performed whereby the same independent radiologists will globally assess all time points for a patient in the review period and adjust an overall assessment if necessary. Finally, if the overall assessment for at least one time point for a patient does not agree between the two independent radiologists, a third independent radiologist will adjudicate and identify which radiologist's assessments they agree with and should be used (for the whole patient). The independent reviewers will be blinded to treatment. Further details regarding this process will be provided in the Charter Review developed separately from the SAP. 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 v1.1 assessment conducted by the investigator.

If the independent reviewers cannot identify any target lesion and any non-target lesion at baseline, then there is "no evidence of disease" (NED) at baseline. In this scenario the later overall visit responses based on the BICR can only be PD (if a new lesion is assessed) or NED (if no new lesion is assessed) or NE (if the images are not evaluable).

Of note, the primary efficacy assessment based on the independent central review results will be used as a pivotal study result.

The primary variable in this study is ORR, defined as the proportion of subjects with a best overall response (BOR) of CR or PR (by RECIST v1.1). A BOR is defined as the

best response (in the order of CR, PR, SD, NED, PD, and NE) among all post-baseline disease assessments that occur until progression, or last evaluable assessment in the absence of progression prior to the initiation of subsequent anti-cancer therapy, irrespective of whether or not subjects discontinued the study treatment.

A subsequent anti-cancer therapy is any anti-cancer therapy started on or after the study treatment discontinuation date. For anti-cancer therapies with a start date equal to the study treatment discontinuation date it is assumed that the anti-cancer therapy was started after the study treatment discontinuation.

The hierarchy used to determine BOR is CR>PR>SD>NED>PD:

- A patient is assigned a BOR of CR if the patient has an overall visit response of CR
- A patient is assigned a BOR of PR if the patient has an overall visit response of PR but does not have a BOR of CR
- A patient is assigned a BOR of SD if the patient has an overall visit response of SD at least 35 days after randomisation (6 weeks from randomisation, and also allowing a 7-day visit window) but does not have a BOR of CR or PR
- A patient is assigned a BOR of NED if the patient has an overall visit response of NED at least 35 days after randomisation. This is only applicable for the BICR.
- A patient is assigned a BOR of PD if the patient has an overall visit response of PD but does not have a BOR of CR, PR or SD
- A patient is also assigned a BOR of PD, if the patient dies without evaluable post-baseline tumour assessments and the death occurs ≤ 91 days (i.e. 2×6 weeks + 7 days) after randomisation.
- Otherwise, a patient is assigned a BOR of NE.

Patients with no evidence of disease at baseline cannot have a BOR of CR or PR and therefore cannot be responders. For the BICR based response rate calculations patients with NED at baseline will be included in the denominator and counted as non-responders.

Patients without a post-baseline tumour assessment will be considered to be non-responders. Patients with a BOR of CR or PR will be classified as responders. All other patients will be classified as non-responders.

The primary analysis of ORR will be based on the independent central radiological assessments. The ORR based on the investigator assessments will also be summarised.

The number and percentage of responders and non-responders will be presented. The 95% Pearson-Clopper CI of ORR for each treatment arm will be provided.

These summaries will be based on the ITT population using the independent central radiological assessments. The summaries will be repeated based on the PPS. The summaries will also be repeated using investigator assessments.

EMA Approach

In order to compare the ORRs in the two treatment arms, the difference in ORRs will be calculated as $ORR_{Arm2} - ORR_{Arm1}$ and a 95% Wald asymptotic CI will be given.

With n_{Arm_i} the number of patients randomised to Arm i (Arm1: EU-Avastin and Arm2: FKB238-002) and p_{Arm_i} the proportion of responders in Arm i , the 95% CI will be calculated as follows:

$$(p_{Arm_2} - p_{Arm_1}) \pm z_{\alpha/2} \sqrt{\frac{p_{Arm_1}(1-p_{Arm_1})}{n_{Arm_1}} + \frac{p_{Arm_2}(1-p_{Arm_2})}{n_{Arm_2}}}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution with $\alpha = 0.05$.

If the 95% CI is within the interval $[\pm 0.1221]$, an equivalence between FKB238 and EU-Avastin, with respect to the ORR, is confirmed.

This analysis will be performed using the PPS and the independent central radiological assessments. A sensitivity analysis will be conducted using the investigator assessments.

FDA Approach

In order to compare the ORRs in the two treatment arms, the ratio in ORRs will be calculated as ORR_{Arm_2}/ORR_{Arm_1} and a 90% asymptotic CI will be given.

With n_{Arm_i} the number of patients randomised to Arm i (Arm1: EU-Avastin and Arm2: FKB238-002) and p_{Arm_i} the proportion of responders in Arm i , the 90% CI will be calculated as follows:

$$\left[\frac{p_{Arm_2}}{p_{Arm_1}} \times \exp(-z_{\alpha/2} \sqrt{v}); \frac{p_{Arm_2}}{p_{Arm_1}} \times \exp(z_{\alpha/2} \sqrt{v}) \right]$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution with

$$\alpha = 0.10 \text{ and } v = \left(\frac{1-p_{Arm_1}}{p_{Arm_1} n_{Arm_1}} \right) + \left(\frac{1-p_{Arm_2}}{p_{Arm_2} n_{Arm_2}} \right)$$

If the 90% CI is within the interval $[0.73, 1.38]$, an equivalence between FKB238 and EU-Avastin, with respect to the ORR, is confirmed.

This analysis will be performed using the ITT population and the independent central radiological assessments. A sensitivity analysis will be conducted using the investigator assessments.

Logistic Regression

Logistic regression analyses will be performed (responder/non-responder as independent binary variable) in an exploratory manner to assess the influence of baseline covariates.

In addition to treatment effect, the covariates as defined in Section 4.10.1.2 will be considered. The covariates based on the randomisation stratification factors will be defined using the original randomisation stratification factors. A multiple imputation approach will be used in case any patient from the ITT population has missing data for any of the covariates included in the logistic model.

Multiple logistic regression analysis will be performed including in the model all the mentioned baseline covariates (treatment-covariate interaction terms will not be included) and the treatment arm.

The odds ratio for treatment (Arm2 versus Arm1) when adjusted for all the mentioned baseline covariates will be estimated, including the corresponding 95% Wald CI.

This analysis will be performed using the ITT population and using the independent central radiological assessments. The analysis will be repeated using investigator assessments.

Sensitivity Analyses

A sensitivity analysis will be conducted where ORR will be based on the BOR recorded from randomisation until PD or death whichever occurs first, without excluding assessments performed after start of any subsequent anti-cancer therapy. Another sensitivity analysis will also be performed for a subset of patients with measurable disease at baseline based on the independent central radiological assessments

The number and percentage of responders and non-responders, the 95% Pearson-Clopper CI of ORR for each treatment arm, and analyses as per EMA approach and FDA approach will be reported for these sensitivity analyses.

The sensitivity analyses will only be conducted using the independent central radiological assessments.

4.10.3 Secondary Efficacy Variables

4.10.3.1 Overall Response Rate at Week 19

ORR (by RECIST v1.1) at Week 19 will be defined as the proportion of subjects with a BOR of CR or PR assessed at Week 19. Only tumour assessment performed up until 19 weeks (i.e., week 18 assessment + 7 day assessment window) from randomisation are considered in this analysis. The BOR is defined in Section 4.10.2.

The descriptive analyses of ORR by timepoints will be produced. For Korea submission, ORR at Week 19 will be analysed using the same methods as for the primary endpoint. These analyses will only be conducted using the independent central radiological assessments.

4.10.3.2 Progression-free Survival

The event of interest for PFS is defined as first documented disease progression or death from any cause, whichever occurs first.

Disease progression will be based on tumour assessments according to RECIST v1.1 criteria. The items of the overall response CR, PR, SD and NED will be taken as progression-free whereas PD will denote disease progression.

PFS is defined as the interval from the date of randomisation until the earlier date of the first documentation of disease progression or death due to any reason. Patients without

the event of disease progression or death at data cut-off will be censored. Both date of disease progression and censoring date are defined below.

Date of Disease Progression

In case a patient experiences disease progression event due to an overall response of PD, the date of first disease progression will be defined based on the dates of the corresponding tumour assessment (i.e. TLs assessment, NTLs assessment and new lesions assessment) with an overall response of PD, using calculation rules as follows:

- If a patient has "Overall Response = PD" due to "TLs = PD", then the date of PD is the date of TLs assessment.
- If a patient has "Overall Response = PD" due to "NTLs = PD", then the date of PD is the date of NTLs assessment.
- If a patient has "Overall Response = PD" due to "New lesions = Yes", then the date of PD is the date of new lesions assessment.
- If a patient has "Overall Response = PD" due to more than one of the cases above, then the date of PD is the earliest of the corresponding dates above.

Censoring Date

In case a patient does not progress or die by the time of data cut-off for analysis, the patient will be censored at the date of last evaluable tumour assessment. The censoring date will be calculated as the latest date of TLs assessment and NTLs assessment.

However, if a patient progresses or dies immediately after two or more consecutive missed or non-evaluable tumour assessments (non-evaluable means an Overall Response of NE), the patient will be censored at the time of the last evaluable tumour assessment with an Overall Response of CR, PR, SD or NED prior to the missed or non-evaluable assessments, see the more detailed description in the section on censored progressions and censored deaths below.

If a patient does not have any post-baseline tumour assessment with an Overall Response of CR, PR, SD or NED, or if a patient has no baseline tumour assessments data, the patient will be censored at Day 1 unless the patient dies within 13 weeks (i.e., 2x6 weeks for tumour assessments + 7-day assessment window) of baseline.

PFS will be calculated in units of months as the difference of the two dates (date of disease progression / death (whatever comes first) or censoring date – date of randomisation + 1) divided by 365.25 and multiplied by 12.

Censored progressions and censored deaths after two or more missed visits:

If a patient progresses or dies immediately after two or more consecutive missed or non-evaluable assessments, the patient will be censored at the time of the latest evaluable post-baseline RECIST assessment prior to the missed or NE assessments (or censored at randomisation in case of no evaluable post-baseline assessment prior to the missed/NE assessments). The allowed gap between the last previous evaluable post-baseline RECIST assessment (or randomisation in case of no previous evaluable post-baseline assessment) and the progression or death, for which the progression/death will still be counted as an event, will be defined according to the rules below.

Considering the assessment window of ± 7 days for the post-baseline tumour assessments the allowed gap will be computed as

$$\text{Allowed gap} = 2 \times 6 \text{ weeks} + 1 \text{ week} = 13 \text{ weeks} = 91 \text{ days}$$

(directly after randomisation)

or

$$\text{Allowed gap} = \text{Sum of planned time intervals for next two assessments} + 2 \text{ weeks}$$

(after a previous evaluable post-BL assessment).

This allows one missing assessment between the previous assessment and the progression or death and considers the allowed visit window of ± 7 days.

The allowed gap after a last previous evaluable post-BL assessment depends on the current assessment schedule of the patient, which can be either “every 6 weeks”, “every 9 weeks”, “every 12 weeks”, or the schedule may just change from every 6 to every 9 weeks, from every 6 to every 12 weeks or from every 9 to every 12 weeks, resulting in the following allowed gaps:

Allowed gap = 13 weeks	(no previous evaluable post-BL assessment)
Allowed gap = 14 weeks	(6 + 6 + 2 weeks, every 6 weeks schedule)
Allowed gap = 17 weeks	(6 + 9 + 2 weeks, schedule changing from every 6 to every 9 weeks)
Allowed gap = 20 weeks	(9 + 9 + 2 weeks, every 9 weeks schedule)
Allowed gap = 20 weeks	(6 + 12 + 2 weeks, schedule changing from every 6 to every 12 weeks)
Allowed gap = 23 weeks	(9 + 12 + 2 weeks, schedule changing from every 9 to every 12 weeks)
Allowed gap = 26 weeks	(12 + 12 + 2 weeks, every 12 weeks schedule)

E.g. if a progression of a patient still on treatment with study drug was assessed exactly 34 weeks after randomisation and the last previous evaluable post-BL assessment was done exactly 17 weeks after randomisation (e.g. with a response “SD”), then the progression is still counted as an event, since the two assessments are within the allowed time windows of the planned assessments at Weeks 18 and 33, so that only the Week 24 assessment is missing (patient changing from every 6 to every 9 weeks).

If the last previous evaluable post-BL assessment was outside the planned visit windows of ± 7 days, then the allowed gap for the latest earlier planned assessment will be used, e.g. if the last previous evaluable post-BL assessment prior to a progression/death was done 16 weeks after randomisation, then an allowed gap of 14 weeks would be used.

See the programming specifications for the detailed definition of the allowed gaps, depending on the last previous evaluable post-BL assessment day and (for patient who discontinued study drug not due to a PD) the study treatment discontinuation day.

PFS will be summarised using Kaplan-Meier estimates of the quartiles (25% quartile, median, 75% quartile) for each treatment arm and 95% CIs for the medians will be calculated.

Additionally, PFS will be presented graphically by the means of Kaplan-Meier curves for each treatment arm.

These analyses will be based on the ITT population using the independent central radiological assessments. Sensitivity analyses will be performed based on the PPS. Analyses will also be repeated using the investigator assessments. However Kaplan-Meier curves will only be produced using the independent central radiological assessments.

Cox Regression

Cox regression analysis will be performed in an exploratory manner to assess the influence of baseline covariates on PFS.

In addition to treatment effect, the covariates as defined in Section 4.10.1.2 will be considered. The covariates based on the randomisation stratification factors will be defined using the original randomisation stratification factors.

Multiple Cox regression analysis will be performed, including in the model all the mentioned baseline covariates (treatment-covariate interaction terms will not be included) and the treatment arm. The hazard ratio for treatment (Arm2 versus Arm1) when adjusted for all the mentioned baseline covariates will be estimated, including the corresponding 95% Wald CI. The assumption of proportionality will be assessed. Proportional hazards will be tested by examining plots of complementary log-log (event times) versus log (time).

In all Cox regression analyses, ties in failure times will be handled using the approximate likelihood of Efron [1].

The Cox regression analysis will be based on the ITT population using the independent central radiological assessments.

Of note multiple imputation approach will be used in case any patient from the ITT population has missing data for any of the covariates included in the Cox model.

Sensitivity Analysis

A sensitivity analysis will be performed to account for the potential effect of investigator's decision to introduce any subsequent anti-cancer therapy before occurrence of the PFS event.

In this analysis, only tumour assessments prior to the start of any subsequent anti-cancer therapy will be considered. Rules to define event or censoring date (see Section 4.10.3.2) will be applied using the set of tumour assessments defined here above.

In this sensitivity analysis, progression-free survival will be analyzed using the same methods as for core PFS as described above using Kaplan-Meier method and Cox regression model.

The sensitivity analysis will be performed using the ITT population and using the independent central radiological assessments.

4.10.3.3 Overall Survival

OS will be evaluated in this study as a secondary endpoint.

The *event* of interest is defined as death from any cause.

OS is defined as the interval from date of randomisation until the date of death due to any cause. Patients without the event of death at data cut-off will be censored at the last date when known to be alive in the study, defined as the maximum of the following:

- date of randomisation
- AE start date, AE stop date, date of AE grade changes
- admission and discharge date of hospitalisation
- date of last visit
- date of last study treatment (IP and/or combination drugs)
- laboratory test dates
- dates related to assessment of potential Hy's law events
- date of vital signs assessments
- date of ECOG PS assessments
- date of ECG assessments
- date of left ventricular ejection fraction (LVEF) assessment
- date of TLs, NTLs and new lesions assessments
- start and stop dates of concomitant medications
- start and stop dates of subsequent anti-cancer treatment
- date of last ADA sample collection
- date of last sample collection for serum drug concentration
- date of withdrawal of consent
- date patient last confirmed to be alive
- date of study completion

In case death info is provided for a patient but the date of death is missing or partially missing, imputation rules will be implemented using a worst case scenario approach.

The time of OS will be calculated in units of months as the difference of the two dates (date of death or censoring date – date of randomisation + 1) divided by 365.25 and multiplied by 12.

OS will be evaluated using same approach as for PFS.

These analyses will be performed using the ITT population and repeated based on the PPS.

A Kaplan-Meier plot of time to censoring will be provided based on the ITT population with the censoring indicator reversed [2]. A corresponding summary table will be provided based on the ITT population.

Cox Regression

Cox regression analysis will be performed in an exploratory manner to assess the influence of baseline covariates on OS. Same approach as described in Section 4.10.3.2 will be used.

The Cox regression analysis will be based on the ITT population.

4.10.3.4 Duration of Response

The DOR will be evaluated in this study as another secondary efficacy endpoint.

Only the patients defined as responders in the primary analysis of ORR, as outlined in Section 4.10.2, will be taken into account for the analysis of DOR.

For this specific endpoint, study day 1 will be defined as the date of first documented objective response.

As for PFS, the *event* of interest is defined as first documented disease progression or death due to any reason, whichever occurs first (see Section 4.10.3.2).

DOR is defined as the interval from the first documented response (as defined per RECIST v1.1) until the earlier date of the first documented disease progression or death due to any reason. The date of first documented response will be taken as the date of the first tumour assessment with an overall visit response of CR or PR.

The PFS censoring rules will be applied (see Section 4.10.3.2).

DOR will be calculated in units of months as the difference of the two dates (date of disease progression / death (whatever comes first) or censoring date – date of first documented response + 1) divided by 365.25 multiplied by 12.

DOR will be evaluated using same approach as used for PFS.

As the patients defined as responders might not be comparable between treatment arms, any statistical comparisons of DOR between treatment arms will be given in a purely exploratory manner and should be interpreted with caution.

These analyses will be based on the ITT population using the independent central radiological assessments. Sensitivity analyses will be performed based on the PPS. Analyses will also be repeated using the investigator assessments. However Kaplan-Meier curves will only be produced using the independent central radiological assessments.

Cox Regression

Cox regression analysis will be performed in an exploratory manner to assess the influence of baseline covariates on DOR. Same approach as described in Section 4.10.3.2 will be used.

The Cox regression analysis will be based on the subset of responders (as defined in Section 4.10.2) in the ITT population and using the independent central radiological assessments.

Sensitivity Analysis

A sensitivity analysis will be performed on DOR using the same censoring rule as for PFS sensitivity analysis as described in Section 4.10.3.2, i.e., progression event will not be counted if it occurs after initiation of other anticancer therapy or occurs immediately after ≥ 2 consecutively missed disease assessments.

Only the patients defined as responders in the primary analysis of ORR, as outlined in Section 4.10.2, will be taken into account for this sensitivity analysis of DOR.

This analysis will be based on the ITT population using the independent central radiological assessments

Similar summary as the one produced for the main analysis of DOR will be produced for this sensitivity analysis.

4.10.3.5 Disease Control Rate

DCR will be defined using the same approach as for the definition of ORR (see Section 4.10.2). The difference to ORR is that in addition to patients with a BOR of CR, PR, patients with a BOR of SD or NED will also be classified as responders.

The DCR will be defined as the proportion of patients defined as responders.

The main analysis of DCR will be based on the independent central radiological assessments. The DCR based on the investigator assessments will also be summarised.

The number and percentage of responders and non-responders and the 95% Pearson-Clopper CI of DCR for each treatment arm will be provided. The odds ratio for treatment (Arm2 versus Arm1) and the corresponding 95% Wald CI will be produced based on a logistic regression analysis of DCR using the same approach as described in Section 4.10.2.

These summaries will be based on the ITT population using the independent central radiological assessments. The summaries will be repeated based on the PPS. The summaries will also be repeated using the investigator assessments.

4.10.4 Tumour and Response Assessments

All tumour assessments data from investigator assessments as well as data from the independent central radiological assessments will be itemised in by-patient-by-visit listings.

4.10.5 Efficacy Endpoints Summary

Efficacy endpoints data will be itemised in by-patient listings.

For endpoints related to tumour assessments, both values based on the independent central radiological assessments and values based on the investigator assessments will be included.

4.11 Safety Evaluation

All safety summaries and analyses will be based on the Safety Population as defined in Section 4.5. Baseline is defined as the last available value prior to the first dose of study treatment, unless specified.

4.11.1 Extent of Exposure

Upon randomisation, patients will enter the Study Treatment Period. Following start of study treatment, patients will attend for study visits every 3 weeks as long as they are receiving study treatment. The combination drugs (paclitaxel + carboplatin) will be administered on Day 1 of each 21-day cycle for at least 4, and no more than 6 cycles. The number of cycles is determined by patients' need and the investigator's assessment. FKB238 or EU-Avastin (IP) will also be administered on Day 1 of each 21-day cycle until objective PD or other criteria for treatment discontinuation are met.

All IP will be administered at the study site by study personnel. Receipt, accountability and return of IP kits will be recorded in ClinPhone@RTSM.

Two types of summaries will be produced to describe exposure to IP and to combination drugs, as follows:

- Summary of study treatment exposure, including notably duration of exposure and relative dose intensity
- Summary of delay and dose modifications

Further details to these summaries are provided below.

4.11.1.1 Exposure and Dose Intensity

Exposure to IP

Exposure to IP will be summarised by treatment arm as follows:

- Descriptive statistics of number of IP cycles received
- Number and percentage of patients per number of IP cycles received
- Descriptive statistics of duration of exposure to IP
- Number and percentage of patients per categories of duration of exposure to IP (≤ 12 weeks, >12 weeks to ≤ 18 weeks, >18 weeks to ≤ 24 weeks, >24 weeks to ≤ 30 weeks, >30 weeks to ≤ 36 weeks, >36 weeks)

The number of IP cycles received will be the number of treatment cycles with an IP administration documented in the eCRF with a positive dose.

The duration of exposure to IP will be calculated in units of weeks as the difference of two dates (date of last IP administration – date of first IP administration + 21) divided by 7. For subjects who die, are discontinued, or lost to follow-up or have the data cut-off prior to the end of the 21 days from last dose, duration of exposure to IP is defined as (the earliest date of death, study discontinuation, last known to be alive and data cut-off – first

IP dosing date + 1) / 7. Only IP administration documented with a positive dose in the eCRF will be considered in this calculation.

Dose Intensity of IP Administration

To further describe IP administration, measurement of dose intensity will be performed. The dose level in mg/kg at each cycle will be calculated based on the total dose in mg of IP administered at a given cycle and the weight measurement in kg used to calculate dosing at the given cycle.

$$Dose\ level\ [mg/kg] = \frac{Total\ dose\ [mg]}{Weight\ [kg]}$$

The cumulative dose level in mg/kg will be defined as the sum of the dose level across all cycles of the study treatment period. The dose intensity in mg/kg/3 weeks will then be calculated based on the cumulative dose level received during the study treatment period and the duration of the study treatment period as follows:

$$Dose\ intensity\ [mg/kg/3\ weeks] = \frac{Cumulative\ dose\ level\ [mg/kg]}{(last\ dosing\ date - first\ dosing\ date + 21)/21}$$

The relative dose intensity will be calculated as the dose intensity divided by the planned dose per cycle (15 mg/kg/3 weeks).

Following summaries will be provided by treatment arm:

- Dose intensity (mg/kg/3 weeks)
- Relative dose intensity (<60%, 60 to <80%, 80 to <90%, 90 to <100%, 100 to <110%, ≥110%)

Exposure to Combination Drugs

Exposure to combination drugs will be summarised separately for paclitaxel and carboplatin. Summaries for each drug will be produced as follows:

- Descriptive statistics of number of drug cycles received
- Number and percentage of patients per number of drug cycles received
- Descriptive statistics of duration of exposure to drug

The number of drug cycles received will be the number of treatment cycles with a drug administration documented in the eCRF with a positive dose.

The duration of exposure to drug will be calculated in units of weeks as follows: (last combination drug dosing date – first combination drug dosing date + 21) / 7. For subjects who die, are discontinued or lost to follow-up or have the data cut-off prior to the end of the 21 days from last dose of combination drug, the duration of exposure to drug is defined as (the earliest of the date of death, study discontinuation date, last date known to be alive and data cut-off date – first combination drug dosing date + 1) / 7. Only drug administration documented with a positive dose in the eCRF will be considered in this calculation.

Dose Intensity of Paclitaxel Administration

To further describe administration of paclitaxel, measurement of dose intensity will be performed. The dose level in mg/m² at each cycle will be calculated based on the total dose in mg of paclitaxel administered at a given cycle and the body surface area (BSA) in m² at the given cycle.

$$\text{Dose level [mg/m}^2\text{]} = \frac{\text{Total dose [mg]}}{\text{BSA [m}^2\text{]}}$$

The BSA at a given cycle will be calculated using the most recent height measurement on or prior to the day of dosing and the weight measurement used for dosing at the given cycle according to the Dubois and Dubois formula: $\text{BSA[m}^2\text{]} = 0.20247 \times \text{Height[m]}^{0.725} \times \text{Weight[kg]}^{0.425}$.

The cumulative dose level will be defined as the sum of the dose level across all cycles of the study treatment period. The dose intensity in mg/m²/3 weeks will then be calculated based on the cumulative dose level received during the study treatment period and the duration of the study treatment period as follows:

$$\text{Dose intensity [mg/m}^2\text{/3 weeks]} = \frac{\text{Cumulative dose level [mg/m}^2\text{]}}{(\text{last dosing date} - \text{first dosing date} + 21)/21}$$

The relative dose intensity will be calculated as the dose intensity divided by the planned dose per cycle (200 mg/m²/3 weeks).

Following summaries will be provided:

- Dose intensity (mg/ m²/3 weeks)
- Relative dose intensity (<60%, 60 to <80%, 80 to <90%, 90 to <100%, 100 to <110%, ≥110%)

Dose Intensity of Carboplatin Administration

To further describe administration of carboplatin administration, measurement of dose intensity will be performed. The area under curve (AUC) dose level at each cycle will be calculated based on the total dose in mg of carboplatin administered at a given cycle and the glomerular filtration rate (GFR) in ml/min at the given cycle using the Calvert formula.

$$\text{AUC Dose level} = \frac{\text{Total dose [mg]}}{\text{GFR [ml/min]} + 25}$$

The GFR at a given cycle will be calculated based on age at Screening, the most recent serum creatinine measurement on or prior to the day of dosing and the weight measurement used for dosing at the given cycle using the Cockcroft-Gault formula.

$$\text{GFR[mL/min]} = \frac{(140 - \text{age[years]}) \times \text{weight[kg]} \times (0.85 \text{ if female or } 1.0 \text{ if male})}{72 \times \text{serum creatinine[mg/dL]}}$$

The cumulative AUC dose level will be defined as the sum of the AUC dose levels across all cycles of the study treatment period. The dose intensity per 3 weeks will then be calculated based on the cumulative AUC dose level received during the study treatment period and the duration of the study treatment period as follows:

$$\text{Dose intensity (AUC/3 weeks)} = \frac{\text{Cumulative AUC}}{(\text{last dosing date} - \text{first dosing date} + 21)/21}$$

The relative dose intensity will be calculated as the dose intensity divided by the planned dose per cycle (AUC 6) of 3 weeks.

Following summaries will be provided:

- Dose intensity
- Relative dose intensity (<60%, 60 to <80%, 80 to <90%, 90 to <100%, 100 to <110%, ≥110%)

4.11.1.2 Delay and Dose Modification

Summaries of IP interruptions and delay will be produced. Similarly, summaries of combination drugs dose reduction, interruptions and delay will be produced.

A by-patient listing of exposure data will be provided for both IP and combination drugs.

4.11.2 Adverse Events

AEs will be collected from time of signature of informed consent, throughout the treatment period and up to and including the 30-days after the last dose of study treatment. All ongoing and any new AEs/SAEs identified during the 30 calendar days after last dose of study treatment must be followed to resolution. SAEs will be recorded from the time of informed consent.

AEs will be coded using the latest MedDRA version. Missing coding terms will be listed and summarised as "Not coded".

Intensity of AEs will be assessed according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 4 grading system.

AEs that start after the date of first administration of study treatment (IP and/or combination drugs) and up to 30 days after the last administration of study treatment (IP and/or combination drugs), or whose severity worsen on or after the date of first administration of study treatment (IP and/or combination drugs) and up to 30 days after the last administration of study treatment (IP and/or combination drugs) will be defined as treatment-emergent AEs (TEAEs).

AEs with onset on the date of first administration of study treatment (IP and/or combination drugs) will not be considered as TEAEs if it is recorded in the eCRF that onset of the AE is prior to first administration of study treatment (IP and/or combination drugs) and if the severity of the AE does not worsen up to 30 days after the last administration of study treatment (IP and/or combination drugs). Otherwise the AE will be considered as a TEAE.

Where dates are missing or partially missing, AEs will be assumed to be TEAEs, unless there is clear evidence (through comparison of partial dates) to suggest that the AE started prior to the first dose of study treatment (IP and/or combination drugs) or more than 30 days after the last dose of study treatment (IP and/or combination drugs).

An AE with missing causal relationship to study treatment (IP and/or combination drugs) will be considered as 'causally related' to study treatment.

AEs will be classified by SOC and PT according to MedDRA. Summary tables will display the number and percentage of patients experiencing the TEAEs and number of TEAEs. In this style of output, "number of TEAEs" will include all occurrences of a TEAE including repeat occurrences in individual patients, while "number of patients" will count each subject only once.

If a patient experiences more than 1 episode of a particular AE, the patient will be counted only once for that event. If a patient has more than 1 AE that is coded to the same PT, the patient will be counted only once for that PT. Similarly, if a patient has more than 1 AE within a SOC, the patient will be counted only once in that SOC.

Overall summaries of AEs will be produced to present the number of events and number and percentage of patients with:

- Any TEAE
- Any TEAE with CTCAE grade 3 or higher
- Any TEAE causally related to IP with CTCAE grade 3 or higher
- Any TEAE causally related to any combination drugs with CTCAE grade 3 or higher
- Any TEAE causally related to study treatment (IP and/or combination drugs) with CTCAE grade 3 or higher
- Any TEAE leading to discontinuation of IP
- Any TEAE leading to discontinuation of any combination drugs
- Any TEAE leading to discontinuation of study treatment
- Any TEAE causally related to IP and leading to discontinuation of IP
- Any TEAE causally related to any combination drugs and leading to discontinuation of any combination drugs
- Any TEAE causally related to study treatment and leading to discontinuation of study treatment
- Any TEAE causally related to IP
- Any TEAE causally related to any combination drugs
- Any TEAE causally related to study treatment
- Any serious TEAE
- Any serious TEAE causally related to IP
- Any serious TEAE causally related to any combination drugs
- Any serious TEAE causally related to study treatment
- Any TEAE leading to death
- Any TEAE causally related to IP and leading to death
- Any TEAE causally related to any combination drugs and leading to death
- Any TEAE causally related to study treatment and leading to death

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

- All TEAEs
- All TEAEs with CTCAE grade 3 or higher
- All TEAEs by maximum CTCAE grade
- All TEAEs causally related to IP with CTCAE grade 3 or higher
- All TEAEs causally related to study treatment with CTCAE grade 3 or higher
- All TEAEs leading to discontinuation of IP
- All TEAEs leading to discontinuation of study treatment
- All TEAEs causally related to IP and leading to discontinuation of IP
- All TEAEs causally related to study treatment and leading to discontinuation of study treatment
- All TEAEs causally related to IP
- All TEAEs causally related to study treatment
- All serious TEAEs
- All serious TEAEs causally related to IP
- All serious TEAEs causally related to study treatment
- All TEAEs leading to death
- All TEAEs causally related to IP and leading to death
- All TEAEs causally related to study treatment and leading to death

In addition, a truncated AE table of most common AEs, showing all events that occur in at least 5% of patients in any treatment arm will be summarised by PT. The threshold of 5% of patients in any treatment arm is based on the PT level.

AEs will be assigned CTCAE grades and summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, SOC, PT and actual treatment group. Fluctuations observed in CTCAE grades during study will be listed (where collected). When assigning a maximum CTCAE grade to a TEAE, only grade changes occurring after the date of first administration of study treatment (IP and/or combination drugs) and up to 30 days after the last administration of study treatment (IP and/or combination drugs) will be considered.

A summary of time to onset of first AE will be provided overall and by SOC and PT. Only TEAEs will be considered in this summary. The time to onset of first AE will be defined in units of days as the difference of the two dates (date of AE onset – date of first administration of study treatment (IP and/or combination drugs) + 1). The earliest onset date of a TEAE overall and by SOC and PT will be considered as applicable.

The number and percentage of patients with the following TEAEs will be summarised by subgroups defined based on ADA levels (Section 4.12.1):

- Any TEAE
- Any TEAE causally related to IP
- Any TEAE of CTCAE grade 3 to 4
- Any TEAE of CTCAE grade 3 to 4 causally related to IP
- Any TEAE leading to death
- Any serious TEAE
- Any serious TEAE leading to discontinuation of IP
- Any TEAE of special interest (see Section 4.11.3)
- Any TEAE of special interest (hypersensitivity reactions, infusion reactions)

Following AEs data listings will be provided:

- A by-patient listing of all AEs (including non-treatment-emergent events)
- A by-patient listing of all serious TEAEs
- A by-patient listing of all TEAEs causally related to IP
- A by-patient listing of all TEAEs with CTCAE grade 3 or higher
- A by-patient listing of all TEAEs leading to discontinuation of IP
- A by-patient listing of all TEAEs leading to discontinuation of study treatment
- A by-patient listing of all TEAEs leading to death

4.11.3 Adverse Events of Special Interest

A list of MedDRA SOC and PT will be provided by the Sponsor to identify the AEs of special interest in the clinical database.

Following summary tables will be repeated for the subset of TEAEs of special interest:

- All TEAEs of special interest
- All TEAEs of special interest by maximum CTCAE grade
- All TEAEs of special interest with CTCAE grade 3 or higher
- All TEAEs of special interest leading to discontinuation of IP
- All TEAEs of special interest leading to discontinuation of study treatment
- All TEAEs of special interest causally related to IP
- All TEAEs of special interest causally related to study treatment
- All TEAEs of special interest leading to death
- All serious TEAEs of special interest
- All TEAEs of special interest with outcome of recovered/resolved or recovered/resolved with sequelae

The summaries will present number and percentage of patients experiencing at least one TEAEs of special interest by group term and each PT within it.

A by-patient listing of all AEs of special interest (including non-treatment-emergent events) will be provided.

4.11.4 Deaths and Other Significant Events

Deaths

Deaths events will be reported as on-treatment deaths if last study treatment administration occurs on the date of death.

A summary of death cases based on the ITT population will be provided presenting number and percentage of patients with death event by period (overall, on-treatment, within 30 days of last study treatment dose, more than 30 days after last study treatment dose) and by relationship to disease under investigation.

An additional summary of death cases based on the Safety population will also be provided presenting number and percentage of patients with death event by period (on-treatment, within 30 days of last study treatment dose, more than 30 days after last study treatment dose) and by following categories:

- Death related to disease under investigation only
- AE with outcome of death only
- Death related to disease under investigation and AE with outcome of death
- Other

A by-patient listing of death events will be provided.

Overdose

A summary of TEAEs associated with overdose by SOC and PT will be provided.

A by-patient listing of overdose events will be provided.

Hy's Law

A by-patient listing of potential Hy's Law cases will be provided.

In addition a by-patient-by-visit listing will be provided including all patients who meets any of the following identification criteria (in isolation or in combination) after start of study treatment:

- Alanine transaminase (ALT) ≥ 3 x Upper Limit of Normal (ULN)
- Aspartate transaminase (AST) ≥ 3 x ULN
- Total bilirubin ≥ 2 x ULN

The listing will include centre, patient identifier, age, gender, race, visit, start date of study treatment (IP and combination drugs), visit, sample collection date, ALT, AST and total bilirubin measurements along with the corresponding criteria flags (≥ 3 x ULN or ≥ 2 x ULN as applicable). Visits at which any of the identification criteria are met will be flagged. Patients with ALT or AST ≥ 3 x ULN together with total bilirubin ≥ 2 x ULN will be flagged.

4.11.5 Clinical Laboratory Evaluation

Blood and urine samples for determination of clinical chemistry, haematology, urinalysis and coagulation will be taken pre-dose at the times indicated in the Schedule of Assessments table from the protocol.

Additional safety laboratory samples may be collected if clinically indicated at the discretion of the investigator. The date and results (values, units and reference ranges) will be recorded on the appropriate eCRF.

The laboratory variables to be measured are summarised in the laboratory variables table from the protocol.

In this study, laboratory parameters measurements will be performed by local laboratories. Measurements will be converted to International System of Units (SI) units.

A summary of each laboratory parameter at Baseline and each post-Baseline time point will be provided for all clinical chemistry, haematology and coagulation laboratory parameters listed in the laboratory variables table from the protocol. A summary of change from Baseline and the percent change from Baseline will be provided for all clinical chemistry, haematology and coagulation laboratory parameters listed in the laboratory variables table from the protocol. Values up to the treatment discontinuation date will be considered in the summary tables.

For the laboratory parameters with available NCI-CTCAE grading criteria, shifts from baseline to the worst grade from start of study treatment up to the treatment discontinuation date will be summarised.

By-patient listings of all laboratory data will be provided with abnormal values highlighted. NCI-CTCAE grade will also be included in the listings.

A by-patient listing of pregnancy test data will be provided. In the event of any pregnancy, pregnancy report data will also be listed.

4.11.6 Vital Signs, Physical Findings and Other Observations Related to Safety

Vital Signs

Vital signs will be assessed according to the Schedule of Assessments table from the protocol and will include systolic and diastolic blood pressure (mmHg), pulse rate (beats per minute), and body temperature (°C). Assessments will be performed pre-dose, at the visits as shown in the Schedule of Assessments table from the protocol, and on occurrence of any cardiac AE.

Summaries of vital signs assessments will be provided for baseline values and post-baseline values up to the latest of the treatment discontinuation date and the last dose of study treatment + 37 days, to consider the 30-days follow-up visit planned 30± 7 days after the last dose of study treatment. Descriptive statistics will be reported for systolic blood pressure, diastolic blood pressure, pulse rate, and body temperature. Actual values and change from baseline will be reported by parameter and time point.

A by-patient listing of vital signs will be provided.

ECG

ECGs are required within 7 days prior to starting study treatment, on Day 1 of Cycle 1, 2 hours within pre-dose and post-dose, and when clinically indicated and at the Follow-up visit after the patient has discontinued the study treatment.

Overall evaluation of ECG will be summarised by time point presenting number and percentage of patients by overall evaluation (normal, abnormal, borderline) and by relevance of abnormality (clinically significant, not clinically significant).

In addition, a shift table will be produced comparing the baseline overall evaluation of ECG and the last on treatment ECG. The shift table will use the following categories of overall evaluation of ECG (normal, abnormal - not clinically significant, abnormal - clinically significant, borderline, not done).

Baseline values and post-baseline values up to the latest of the treatment discontinuation date and the last dose of study treatment + 37 days will be considered in the summary tables, taking into account the ± 7 days window for the 30-days follow-up visit.

A by-patient listing of ECG data will be provided.

LVEF

LVEF by cardiac ultrasound will be performed at Screening and at study treatment discontinuation visit.

A summary by time point of LVEF data will be provided.

In addition, number of subjects by treatment arm and overall in following categories will be displayed:

- Baseline LVEF measurement available
- Any post-baseline LVEF measurement available
- Any occurrence at a post-baseline LVEF assessment of a LVEF decrease of ≥ 10 percentage points along with a LVEF $< 50\%$
- Any occurrence at a post-baseline LVEF assessment of a LVEF decrease of ≥ 15 percentage points along with a LVEF $\geq 50\%$.

Post-baseline values up to the treatment discontinuation date will be considered for the above summaries.

A by-patient listing of LVEF data will be provided.

Physical Examination

Physical examination will be performed as described in the Schedule of Assessments table from the protocol.

A shift table from baseline to post-baseline physical examination will be produced. The table will summarise by body system the baseline status versus worst status from start of study treatment up to the latest of the treatment discontinuation date and the last dose of study treatment + 37 days, considering the ± 7 days time window for the 30-days follow-up visit.

A by-patient listing of abnormalities identified from physical examination will be provided.

ECOG PS

A by-patient listing of ECOG PS will be provided.

4.11.7 Safety Monitoring (Independent Data Monitoring Committee)

An Independent Data Monitoring Committee (IDMC) will be involved in the conduct of this study. The IDMC has the responsibility for monitoring the progress of the clinical study and the safety of the study participants. The IDMC will perform review of safety data and study conduct during the course of the clinical trial, as defined in the IDMC Charter for this clinical trial. The IDMC will not stop the trial for efficacy. The memberships of the IDMC and reporting structure are defined in the IDMC Charter.

4.12 Other Analyses

Following sections describe analysis of immunogenicity and PK data. Serum for measurement of the presence of ADAs for FKB238 and EU-Avastin will be collected according to the schedule in Table 3 and will be analysed using validated methods. A tiered analysis approach will be followed where samples positive for ADA at screening step will be confirmed for ADA presence. The confirmed positive samples will then be assessed for ADA titre and the presence of neutralizing antibodies (nAb).

Table 3 Schedule of Pharmacokinetic and Antidrug Antibody Sampling

	Cycle 1 Day 1	Cycle 2 Day 1	Cycle 4 Day 1	Cycle 6 Day 1	Study Treatment Discontinuation Visit	Follow-up
Pre-dose	PK/ADA	PK/ADA	PK/ADA	PK/ADA	PK/ADA	PK/ADA ¹
Immediately after completion of IP infusion	PK		PK			

¹ Follow-up samples will be taken every 12 weeks (up to 1 year [\pm 14 days] after randomisation) until death, or the patient is lost to follow-up, whichever occurs first.

4.12.1 Immunogenicity

ADA and nAb results will be listed and summarised by treatment arm and timepoint using descriptive statistics based on the ADA evaluable population.

Of note, the following categories will be defined:

- ADA prevalence: ADA positive at any time point including before first study treatment administration.
- Treatment-boosted ADA: Treatment-boosted ADA is defined as baseline positive ADA titer that was boosted to a 4 fold or higher level following drug administration
- Persistent positive ADA: ADA positive at ≥ 2 post-baseline assessments (with ≥ 16 weeks between first and last positive) or positive at last post-baseline assessment
- Transient positive ADA: At least one post-baseline ADA positive assessment and not fulfilling the conditions of persistent positive

Relevant efficacy and safety summaries might be produced for subgroups defined based on ADA levels.

All immunogenicity data will be listed by treatment/subject/timepoint.

4.12.2 Pharmacokinetics

Trough (C_{trough}) and peak (C_{max}) serum concentrations will be compared by treatment arms and timepoints and descriptive statistics will be provided. The impact of ADA on C_{trough} and C_{max} will be assessed descriptively. The pre-dose serum concentrations at cycles 2, 4 and 6 will be considered as C_{trough} values and the post-dose serum concentrations at cycles 1 and 4 as C_{max} values.

Serum concentrations of FKB238 and Avastin will be listed and summarised using the PK population for each visit at which samples were taken including notably the geometric mean and geometric coefficient of variation (CV).

The pre-dose serum concentrations will be listed but excluded from summary statistics if the subject did not receive IP dose in previous cycle except for cycle 1 (for example, if a subject did not receive cycle 1 day 1 dose, then cycle 2 day 1 pre-dose concentration will be excluded from summary statistics).

The post-dose serum concentrations will be listed but excluded from summary statistics if the subject did not receive IP dose in this cycle.

Serum concentration values below the lower limit of quantification (BLQ) will be handled as follows in the computation of summary statistics (within each timepoint and treatment group):

- If at most 50% of values are below the lower limit of quantification (LLOQ), then BLQ values will be imputed with a value of LLOQ in the computation of summary statistics.
- If >50% but not all values are BLQ, then the mean, standard deviation, CV, geometric mean and geometric CV will be shown as “NC” (not calculable), whereas minimum and median will be shown as “BLQ”.
- If all values are BLQ, then the mean, minimum, median, maximum and geometric mean will be shown as “BLQ”, whereas standard deviation, CV and geometric CV will be shown as “NC”.

In addition, summary statistics of C_{trough} and C_{max} will also be presented by subgroups defined based on ADA levels for each visit at which samples were taken using the PK population.

Time-course plot of pre-dose serum concentrations by treatment arm will be presented using the PK population. Plots of individual pre-dose serum concentrations will also be provided by treatment arm/timepoints and by subgroups defined based on ADA levels using the PK population.

A by-patient/treatment/timepoint listing of PK data will be provided.

4.13 Determination of Sample Size

Assuming a dropout rate of 10%, it is anticipated that approximately 730 patients will be randomised into the study in a 1:1 ratio (365 patients in the FKB238 group and 365 patients in the Avastin group) in order to have a total of 656 patients who complete study treatment. Sample size was determined to meet both the EMA and FDA requirements, which differ in several aspects.

To fulfil the EMA requirements, a meta-analysis of available randomised clinical studies of Avastin demonstrated that the risk-difference for the ORR for the control arm compared to the Avastin treatment arms was calculated to be 0.1938 (80% CI: [0.1564, 0.2312]). Based on the result of the meta-analysis, an equivalence margin for the risk-difference was determined to be 0.1221, which preserves 22% of the treatment effect characterised by the lower 80% CI for the risk-difference of ORR.

The NCSS PASS 2005 statistical software was used for the sample size calculation. With the equivalence margin for the risk-difference of 0.1221, an expected response rate of 35% in both treatment arms, the study design employing a two one-sided test (TOST) procedure, and an overall Type I error rate of 2.5%, a sample size of 656 patients (328 per group) was calculated to provide 80% power to demonstrate that the 95% CI about the risk-difference comparing FKB238 and EU-Avastin falls completely within ± 0.1221 . A total of 730 NS-NSCLC patients (allowing for a dropout rate of 10%) will be randomised either to the FKB238 group or Avastin group in a 1:1 ratio in a parallel group study.

Per the FDA requirements, a meta-analysis was performed on data from relevant clinical trials of Avastin, resulting in an estimate of the risk-ratio of 0.5212 along with a 70% CI,

(0.4775, 0.5689). With lower and upper equivalence margins of 0.73 and 1.38, which preserve 50% treatment effect determined by the lower limit respectively of the 70% CI, an expected response rate of 35% in both treatment arms, a TOST procedure for equivalence, and the overall Type I error rate at 5%, a sample size of 656 patients (328 per group) provides 80% power to demonstrate that the 90% CI for the risk-ratio comparing FKB238 and EU-Avastin is entirely enclosed within 0.73 to 1.38. Accounting for potential 10% dropouts, a total of 730 patients will be randomised.

4.14 Changes in the Conduct of the Study or Planned Analysis

Based on version 5 of the study protocol there are no changes in the conduct of the study or the planned analyses, but an additional analysis population (“ADA evaluable population”) is defined in this SAP and not mentioned in study protocol version 5.

5 REFERENCES

- [1] Efron, B. The Efficiency of Cox’s Likelihood Function for Censored Data, *Journal of the American Statistical Association*, 1977, 72: 557–565.
- [2] Shuster JJ. Median follow-up in clinical trials. *Journal of Clinical Oncology*, 1991, 9(1):191-192.

PAREXEL International Electronic Signature Page

This page is the manifestation of the electronic signature(s) used in compliance with PAREXEL International's electronic signature policies and procedures and in compliance with applicable regulations.

UserName: [REDACTED]
Title: Principal Biostatistician, GDO
Date: Thursday, 07 March 2019, 06:39 PM GMT Standard Time
Meaning: Document contents approved.
