Title: A Randomized Phase 2 Study of AP26113 in Patients with ALK-positive, Non-small Cell Lung Cancer (NSCLC) Previously Treated with Crizotinib

NCT Number: NCT02094573

SAP Approve Date: 25 February 2016

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable (PPD) information or company confidential information (CCI).

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

STATISTICAL ANALYSIS PLAN

STUDY TITLE: A Randomized Phase 2 Study of AP26113 in Patients with ALK-positive, Non-small Cell Lung Cancer (NSCLC) Previously Treated with Crizotinib

PROTOCOL NUMBER:	AP26113-13-201
PROTOCOL DATE:	27 June 2013
PROTOCOL AMENDMENT DATE:	29 July 2014
SAP VERSION:	Version 1.0
DATE:	25 February 2016
SPONSOR:	ARIAD Pharmaceuticals, Inc. 26 Landsdowne Street Cambridge, MA 02139-4234

TABLE OF CONTENTS

TABLE OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Term
AE	adverse event
ALK	anaplastic lymphoma kinase
ALT	alanine aminotransferase
ANC	absolute neutrophil count
ARIAD	ARIAD Pharmaceuticals, Inc.
AST	aspartate aminotransferase
AUC	area under the curve
CBC	complete blood count
C_{max}	maximum plasma concentration
CNS	central nervous system
CR	complete response
CSR	Clinical Study Report
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
ECG	electrocardiogram
ECHO	echocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organisation for Research and Treatment of Cancer
HRQoL	health-related quality-of-life
IRC	independent review committee
ITT	Intention-to-treat
MRI	magnetic resonance imaging
ms	milliseconds
NE	not evaluable
NSCLC	non-small cell lung cancer
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PK	pharmacokinetic(s)
PR	partial response
PRO	patient-reported outcomes
QD	once-daily
QLQ	Quality of Life Questionnaire
QTcF	QT interval corrected (Fridericia)
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SGOT	serum glutamic oxaloacetic transaminase
SGPT	serum glutamic pyruvic transaminase
SRS	stereotactic radiosurgery
TKI	tyrosine kinase inhibitor
ULN	upper limit of normal

1 INTRODUCTION

This statistical analysis plan (SAP) describes the study design and statistical analyses for study protocol AP26113-13-201, entitled "A Randomized Phase 2 Study of AP26113 in Patients with ALK-positive, Non-small Cell Lung Cancer (NSCLC) Previously Treated with Crizotinib" by ARIAD Pharmaceuticals, Inc.

2 STUDY DESIGN

This is a phase 2, open-label, randomized, multicenter, international study. The patient population will include patients with ALK-positive, locally advanced or metastatic non-small cell lung cancer (NSCLC) whose disease has progressed on therapy with crizotinib. An estimated 218 patients will be enrolled at approximately 100 centers. Patient will be randomized 1:1 to receive AP26113 in one of two different dosing regimens. In Arm A, AP26113 will be administered at a dose of 90 mg orally QD, continuously. In Arm B, AP26113 will be administered at a dose of 90 mg orally QD for 7 days, then 180 mg orally QD continuously. A cycle of therapy will comprise 28 days of treatment, regardless of dose. Patients will remain on treatment until they meet one or more criteria for withdrawal as listed in Section 11.9 of the protocol. Patients will be evaluated according to the Schedule of Events in Appendix 1 (Section 11.1 of the protocol).

2.1 Randomization

Patients will be allocated in a 1:1 ratio to each dosing regimen using a randomized block design stratified by the following factors:

1. Presence of brain metastases at baseline (Yes vs. No)
2. Best prior response to crizotinib therapy as assessed by the investigator (CR/PR vs. Any other response or status unknown)

A stratified permuted block randomization design has been determined for this study. The randomization schedule will be created using SAS® procedure PROC PLAN at ARIAD.

2.2 Study Objectives

2.2.1 Primary Objective

The primary objective of the study is to determine the efficacy of AP26113, as evidenced by objective response rate, in patients with ALK-positive locally advanced or metastatic NSCLC whose disease has progressed on therapy with crizotinib. Two dosing regimens will be tested.

2.2.2 Secondary Objective

Secondary objectives for each dosing regimen are as follows:

1. To further characterize the efficacy of AP26113 in patients with ALK-positive, locally advanced or metastatic NSCLC whose disease has progressed on therapy with crizotinib, as shown by disease control rate, time to/duration of response, progression-free survival (PFS), overall survival (OS), and time on treatment
2. To assess intracranial CNS response and PFS, per a modification of RECIST v1.1, in patients who have active brain metastases
3. To assess the safety and tolerability of AP26113 in study patients

4. To measure steady-state plasma levels of AP26113 for use in population PK modeling
5. To assess patient reported symptoms and health-related quality of life (HRQoL) with the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ)-C30 (v.3.0)

2.2.3 Exploratory Objective

CCI [redacted]

2.3 Study Endpoints

2.3.1 Primary Endpoint

The primary endpoint is confirmed ORR, as assessed by the investigator, per RECIST v1.1.

2.3.2 Secondary Endpoints

The following secondary endpoints will be assessed:

1. Confirmed ORR, as assessed by a central independent review committee (IRC), per RECIST v1.1
2. Intracranial CNS response (ORR and PFS, per a modification of RECIST v1.1, in patients who have active brain metastases), as assessed by a central IRC
3. Time to response
4. Duration of response
5. Time on treatment
6. Disease control rate (the percentage of patients with best response of complete response [CR], partial response [PR], or stable disease [SD]), per RECIST v1.1
7. PFS
8. OS
9. Safety and tolerability
10. Steady-state plasma level of AP26113 for use in population PK modeling
11. Patient-reported symptoms of lung cancer and HRQoL scores, assessed with the EORTC QLQ-C30 (v.3.0)

2.3.3 Exploratory Endpoints

CCI

[redacted]

3 GENERAL CONSIDERATIONS

In general, each dosing regimen will be summarized separately. No inferential comparisons across the two regimens will be performed. Baseline values are defined as the last valid values collected during the time interval from the screening visit to the first dose date (time) of the study treatment. In safety analysis, minimum or maximum values collected during this interval may be used for baseline summaries and change from baseline summaries. Descriptive statistics (such as means, medians, standard deviations, and ranges for continuous data, and counts/percentages for categorical data) will be used to summarize patient characteristics, study treatment administration, efficacy, safety, pharmacokinetic parameters, and genetic status of biomarkers. Data will also be displayed graphically, where appropriate. The primary endpoint will be tested at a two-sided alpha level of 0.025 to adjust for the multiplicity of having two treatment arms tested (with two-sided 97.5% confidence intervals). The primary analysis will be conducted when all ongoing patients have completed their Cycle 5 disease assessment. Two-sided confidence intervals for all other parameters to be estimated will be constructed using a significance level of $\alpha = 0.05$. For the purpose of reporting efficacy and safety at specific time points, e.g., 6-month ORR, a month will comprise 30.4375 days. In case that at least one treated patient have been randomized to dosing regimen of 90 mg QD to 180 mg QD but have never been unable to successfully escalated to 180 mg (either during the first 7 days or later), these patients will still be included on all the planned analyses for this regimen and will also be analyzed as a separate group. All the statistical analyses will be performed using SAS Statistical Software.

This SAP has been prepared to support a clinical study report (CSR) that will be included in regulatory submissions intended to achieve marketing approval. CSR analyses, as directed by this document, may also be used for other purposes such as publication of study results, and analyses will be rerun on updated data. Since study AP26113-13-201 will be ongoing at the time this first CSR is written additional analyses as specified in this SAP are anticipated at least annually. For revisions to the CSR any changes and additions to analyses specified here will be detailed in a companion document that will be included as an appendix.

3.1 Sample Size

This is a phase 2, multicenter, randomized, open-label study in patients with ALK-positive, locally advanced or metastatic NSCLC who were previously treated with crizotinib. This study is designed to determine the efficacy in patients treated with daily oral administration of AP26113 at a dose of 90 mg QD continuously or 90 mg QD for 7 days followed by escalation to 180 mg QD continuously. The primary endpoint is ORR assessed by the investigator using RECIST v1.1. The primary analysis of the primary endpoint in the intention-to-treat (ITT) population will be conducted using exact 2-sided 97.5% binomial confidence intervals for each treatment arm. For the purpose of this study, the uninteresting ORR is set at 20% based on a review of the response rates among ALK-positive NSCLC patients primarily in second-line chemotherapy studies and in consultation with regulatory agencies. The alternative ORR is set at

35%. A sample size of at least 218 eligible patients (109 per treatment regimen) will provide approximate 90% power to rule out an uninteresting rate of 20% when the true rate is 35% or higher at two-sided alpha level of 0.025 using exact binomial test. The treatment regimen will be considered to have achieved the primary objective when ORR assessed by the investigator is shown to be significantly higher than 20% at a two-sided alpha level of 0.025 at final analysis for that regimen.

3.2 Analysis Populations

3.2.1 Intention to Treat (ITT) Population

The ITT population includes all patients randomized to each regimen regardless of whether they receive study drug or adhere to the assigned dose.

The primary analyses of efficacy will be based on the ITT population.

3.2.2 Treated Population

The treated population for each regimen includes all patients receiving at least one dose of study treatment.

Safety will be analyzed using the treated population. In the case that more than 3 patients have been randomized and are not treated with study treatment, the primary efficacy analyses will be repeated using the treated population.

3.2.3 Per-protocol Population

The sensitivity analyses of the primary endpoint and selected secondary efficacy endpoints will be performed using the per-protocol population.

The per-protocol population will exclude all patients in the treated population who do not meet key entry criteria, have no measurable disease at baseline, or have no adequate post-baseline radiographic response assessment. To specify, the per-protocol population will be the treated patients who also meet all of the following criteria:

- Histologically or cytologically confirmed locally advanced or metastatic NSCLC
- Confirmed baseline ALK rearrangement by Vysis® FISH test either locally or by central confirmation
- Previously progressed on crizotinib
- At least one measurable a target lesion as assessed by the investigator
- At least two adequate post-baseline radiographic response assessments unless the reason for no post-baseline radiographic response assessment is one of the following:
    - Death
    - Discontinuation due to documented disease progression per RECIST v1.1
    - Discontinuation due to AE

In cases where the investigator and IRC have different opinions regarding whether a patient has measurable disease at baseline, two types of per-protocol populations will be defined, one based on the investigator's assessment, and the other based on IRC's assessment. Analyses of the ORR will be performed for both populations.

3.3 Demographics, Baseline Characteristics, and Patient Disposition

Demographics and baseline characteristics to be summarized using descriptive statistics are listed in Table 1. Continuous prognostic factors will be summarized by means, medians, standard deviations, and ranges; categorical prognostic factors will be summarized by counts/percentages. Other prognostic factors (e.g., by categorizing the continuous prognostic factors and re-categorizing the categorical prognostic factors) may also be included in the analysis.

Table 1 Demographics and Baseline Characteristics

Parameters	Categories for categorical parameters
Age (continuous)	
Age (categorized in reference to age in whole years)	18 – 64, > = 65
	18-49, 50-64, 65-75, and > = 75 years
Gender	Male, Female
Race	White, Black/African American, Asian, Native Hawaiian or other Pacific Islander, Other
Ethnicity	Hispanic or Latino, Not Hispanic or Latino
Geographic region/investigative sites	North America, Europe, Asia-Pacific
Strata as randomized	Brain metastases at baseline/Best response to prior crizotinib of CR or PR, Brain metastases at baseline/Best response to prior crizotinib of any other response status or unknown, No brain metastases at baseline/Best response to prior crizotinib of CR or PR, No brain metastases at baseline/Best response to prior crizotinib of any other response status or unknown
Current strata	Brain metastases at baseline/Best response to prior crizotinib of CR or PR, Brain metastases at baseline/Best response to prior crizotinib of any other response status or unknown, No brain metastases at baseline/Best response to prior crizotinib of CR or PR, No brain metastases at baseline/Best response to prior crizotinib of any other response status or unknown
Stage at diagnosis	IA, IB, IIA, IIB, IIIA, IIIB, IV, Unknown or not staged
Stage at study entry	IIIB, IV, Other
Time since initial diagnosis (continuous)	
Histopathological classification at study entry	Adenosquamous Carcinoma, Squamous, Large cell, Unknown, Other
Metastatic sites at study entry	Lung, brain, liver, bone, and number
Measureable intracranial CNS metastases as assessed by intracranial CNS IRC	Yes, No
Active intracranial CNS metastases at baseline	Yes, No
Cigarette smoking history	Never, Former, Current
Current cigarette smoking amount	Less than 20 per Day, 20 To 60 per Day, More Than 60 per Day broken out by status as current or former smoker
Weight (continuous)	
Height (continuous)	
BMI (continuous)	
ECOG Performance Status	0, 1, 2
ALK mutation method of assessment	FISH-Vysis, FISH-non Vysis, IHC, RT-PCR, Sequencing, Unknown, Other If other, specify
	ALK+ by FISH-Vysis – Yes or No
ALK mutation type	L1152R, G1269A, S1206Y, F1174L, L1196M, D1203N, C1156Y, T1151Tins, G1202R, V1180L, I1171N, Unknown, Other
Overall Best response to prior crizotinib regimen(s)	Complete Response, Partial Response, Stable Disease, Progressive Disease, Unable to Assess, Unknown, Other
Best response to most recent crizotinib regimen	Complete Response, Partial Response, Stable Disease, Progressive Disease, Unable to Assess, Unknown, Other
Reason for stopping most recent crizotinib regimen	Resistant, Intolerant, Unknown, Other
Total duration of crizotinib (continuous)	
Time from the stop date of most recent crizotinib regimen to first dose date (continuous)	

Parameters	Categories for categorical parameters
Most recent systemic therapy	Crizotinib, chemotherapy, other
Prior anticancer therapies	Top 10 therapies used (e.g., pemetrexed, carboplatin) Number of patients who had received any chemotherapy Number of patients who had received any platinum-based therapy Number of patients who had received any other TKI
Any prior radiotherapy	Yes, No
Prior radiotherapy to the brain	Yes, No
Number of prior systemic anticancer therapies/regimens	0, 1, 2, etc.

Patient disposition will be tabulated by reasons leading to study treatment discontinuation including adverse event, death, documented disease progression (RECIST), clinical progressive disease, withdrawal by subject, physician decision, pregnancy, non-compliance with study drug, protocol violation, lost to follow-up, study terminated by sponsor, or other reason.

The list of characteristics described in the table above will also be used as the basis for subgroup analyses as described in Sections 3.4.3 and 3.4.4 below.

3.3.1 Imputation Rules for Missing Initial Cancer Diagnosis Date and Start Date and Stop Date for Selected Prior Anti-Cancer Therapies

In general, a diagnosis date will be imputed first and then used to adjust the imputation of the corresponding prior treatment start date when necessary.

Initial Diagnosis Dates

- If day is missing but month and year are non-missing (UU-MMM-YYYY), impute as earliest of: 01-MMM-YYYY, randomization date.
- If day and month are missing (UU-UUU-YYYY), impute as earliest of: 01-JAN-YYYY, randomization date.
- No imputation for a completely missing date (UU-UUU- UUUU)
- Additional adjustment(s) may be applied depending on ARIAD medical's review on the prior anti-cancer therapies data

Prior Anti-Cancer Therapies Dates

Start Date

- If day is missing but month and year are non-missing (UU-MMM-YYYY), impute as earliest of: 01-MMM-YYYY, randomization date.
- If day and month are missing (UU-UUU-YYYY), impute as earliest of: 01-JAN-YYYY, randomization date.
- No imputation for a completely missing date (UU-UUU- UUUU)
- If an imputed prior treatment start date is before diagnosis date, re-impute as diagnosis date.

Stop Date

- If day is missing but month and year are non-missing (UU-MMM-YYYY), impute as the earliest of

  - 30-MMM-YYYY if month is April, June, September or November, 28-MMM-YYYY if month is February, 31-MMM-YYY otherwise
  - Randomization date
- If day and month are missing (UU-UUU-YYYY), impute as earliest of: 31-DEC-YYYY, randomization date.

If after applying above rules, any prior treatment stop dates are prior to corresponding prior treatment start date, impute as start date.

Duration of a selected prior anti-cancer therapy will be calculated using the following formula: duration = stop date – start date +1. Duration will be calculated with imputed dates when necessary. Total duration will be the sum of the individual durations.

Time since the stop date of a selected prior anti-cancer therapy to the first dose of study treatment will be calculated using the following formula: time since stop date = stop date – first dose date +1. Time since the stop date to the first dose date will be missing in the case of a completely missing stop date.

3.4 Efficacy Analysis

3.4.1 Disease Assessment

Disease assessment by CT or MRI scans will be performed at screening and at 8 week intervals thereafter (on Day 1 [±3 days] of every odd-numbered Cycle) through 15 cycles after the initial dose of AP26113, and every 3 cycles thereafter until disease progression per RECIST v1.1 as assessed by the investigator. More frequent imaging is permitted at any time, if clinically indicated; confirmation of CR or PR should be performed 4 weeks (allowing a minus 3-day time window) or more after initial response. Imaging assessment will also be performed at the End of Treatment visit if more than 4 weeks have passed since the last imaging assessment.

For patients who continue the study treatment beyond documented progressive disease per RECIST v1.1 at the investigator's discretion, imaging will continue with the same assessment schedule. If the patient experiences symptomatic deterioration in the absence of radiologic progression, it is strongly recommended that additional imaging studies be performed to confirm progressive disease.

3.4.2 Definitions of Efficacy Endpoints

The primary endpoint, ORR assessed by the investigator, is defined as the proportion of the patients who are confirmed to have achieved CR or PR, per RECIST v1.1, after the initiation of study treatment in the ITT population. Confirmed responses are those that persist on repeat imaging 4 weeks (allowing a minus 3-day time window) or more after initial response.

Secondary efficacy endpoints for this study are defined as follows:

- Confirmed ORR assessed by IRC is defined as the proportion of the patients who are confirmed to have achieved CR or PR per IRC using RECIST v1.1 after the initiation of study treatment in the ITT population. Confirmed responses are those that persist on repeat imaging 4 weeks (allowing a minus 3-day time window) or more after initial response.
- For randomized patients with active brain metastases at enrollment, intracranial CNS ORR is defined as the proportion of the patients who have achieved CR or PR in the

intracranial CNS per a modification of RECIST v1.1 as evaluated by IRC after the initiation of study treatment.

- For randomized patients with active brain metastases at enrollment, intracranial CNS PFS as evaluated by IRC is defined as the time interval from the date of the first dose of the study treatment until the first date at which intracranial CNS disease progresses, an increase of 20% or more in the sum of diameters of intracranial CNS target lesions, unequivocal progression of non-target lesions, or the appearance of new lesions in the intracranial CNS, is objectively documented by a scan, or death due to any cause, whichever occurs first. Intracranial CNS PFS will be censored according to the scheme specified in Section 3.4.5.3.
- Unless otherwise stated, secondary efficacy endpoints will use investigator assessments with sensitivity analyses performed using the IRC assessments.
- Disease control rate is defined as the proportion of the randomized patients who are confirmed to have achieved CR or PR or have a best overall response as SD for 6 weeks (allowing a minus 3-day time window) or more after initiation of the study drug using RECIST v1.1
- Time to response is defined as the time interval from the date of the first dose of the study treatment until the initial observation of CR or PR for patients with confirmed CR/PR.
- For patients with a confirmed CR/PR, duration of response is defined as the time interval from the time that the measurement criteria are first met for CR/PR (whichever is first recorded) until the first date that the progressive disease is objectively documented or death. Duration of response will be censored according to the scheme specified in Section 3.4.5.3.
- Time on treatment is defined as the time interval from the first dose to the last dose of AP26113. For patients who have not discontinued, time on treatment will be censored as of the last dose of drug.
- PFS is defined as the time interval from the date of the first dose of the study treatment until the first date at which disease progression is objectively documented, or death due to any cause, whichever occurs first, in the ITT population. PFS will be censored according to the scheme specified in Section 3.4.5.3.
- OS is defined as the time interval from the date of the first dose of the study treatment until death due to any cause in the ITT population. OS will be censored on the date of last contact for patients who are still alive.

3.4.3 Primary Efficacy Endpoints Analysis

The best response (CR, PR, SD, or PD) according to RECIST v1.1 will be derived for each patient who receives at least one dose of study treatment. Patients with no measurable disease at baseline or no adequate post-baseline radiographic response assessment will be included as non-responders in the primary efficacy endpoint analysis. The ORR is calculated as the proportion of the randomized patients who are confirmed to have achieved CR or PR after the initiation of study treatment using the following formula: ORR = (#confirmed Investigator-based CR or PR) /(#Randomized patients)*100%. Confirmed responses are those that persist on repeat imaging 4 weeks (allowing a minus 3-day time window) or more after initial response. Exact 2-sided 97.5% confidence intervals for the ORR will be calculated based on the binomial distribution. The primary analysis will be performed on ORR assessed by the investigator among all the randomized patients.

The best response in target lesions (including pathological lymph nodes) per RECIST v1.1 will be calculated as the maximum unsigned decrease (or the minimum increase if no decrease) in percentage in the sum of the longest dimensions of the target lesions at a single assessment as compared with baseline. The best response will be displayed using a waterfall plot.

Sensitivity analysis I will be performed for ORR assessed by the investigator in the ITT population and using all responses (including unconfirmed responses).

Supportive sensitivity analysis II will be performed for confirmed ORR assessed by the investigator in the per-protocol population.

Subgroup analyses of the investigator-based ORR in the ITT population will be performed in all patient groups shown in Table 1 that have at least 10 patients. The difference in ORR between subgroups will be examined by Fisher's exact test or logistic regression.

Relationships between pharmacological findings and both safety and efficacy findings will be analyzed in a separate exposure-response report. Analyses for the exposure-response report will be conducted under an analytical plan specific to that report.

The primary endpoint ORR will also be summarized by study center (small centers will be combined by region). Due to the small number of patients expected in most of the centers, it has been decided that study center will not be used as an additional randomization stratification factor or in logistic regression analyses.

3.4.4 Secondary Efficacy Endpoint Analyses

ORR per IRC will be used to assess the robustness of the primary analysis of the primary endpoint. The IRC-based ORR in the ITT population will be calculated using the following formula: ORR per IRC = (#confirmed IRC-based CR or PR) / (#Randomized patients)*100%. The calculation of the IRC-based ORR in the per-protocol population will be as follows: ORR per IRC = (#confirmed IRC-based CR or PR) / (#Per-protocol patients)*100%. The exact 2-sided 97.5% binomial confidence intervals will be calculated.

Disease control rate will be assessed by the investigator and IRC in the ITT population and the exact 2-sided 95% binomial confidence intervals will be calculated. The calculation formulas of the disease control rate will be as follows: disease control rate per investigator = (#Investigator-based confirmed CR or PR or SD) / (#Randomized patients)*100%; disease control rate per IRC = (# IRC-based confirmed CR or PR or SD) / (#Randomized patients)*100%.

In general, the calculation of time-to-event efficacy endpoints will be as follows:

- PFS = progression/death/censoring date – first dose date + 1
- OS = death/last contact date – first dose date + 1
- time to response = response/censoring date – first dose date +1
- Duration of confirmed response = PFS event/censoring date – date of the first response that was confirmed + 1.
- Duration of response = PFS event/censoring date – date of first response + 1.

Median values and 2-sided 95% confidence intervals will be estimated using Kaplan-Meier (KM) method (Kaplan and Meier, 1958) in the ITT population. The KM-estimated PFS rates and OS rates at 12 and 24 months and the associated 2-sided 95% confidence intervals will be computed. Time to response will be summarized only for confirmed responders and all

responders (including confirmed and un-confirmed responders) using descriptive statistics, and time to both types of response will be estimated by the MK method in which followup for non-responders will be censored. Duration of response will also be summarized with descriptive statistics for patients with confirmed CR or PR and the Kaplan-Meier method in which followup for patients without PFS events will be censored.

An additional IRC assessment will be performed to assess efficacy endpoints in the intracranial CNS in randomized patients with active brain metastases assessed by MRI at enrollment (see Imaging Services Review Charter). Intracranial CNS ORR assessed by the IRC and the exact 2-sided 95% binomial confidence intervals will be calculated; median intracranial CNS PFS will be estimated using the Kaplan-Meier method. The calculation formulas for intracranial CNS ORR will be as follows: intracranial CNS ORR per IRC = (#IRC-assessed confirmed intracranial CNS CR or PR) / (#Randomized patients with active brain metastasis)*100%.

Subgroup analyses of selected secondary efficacy endpoints will be performed in the same manner as described for the primary endpoint above.

The correlation of ALK and other biomarkers with time-to-event endpoints will be evaluated using Cox proportional hazard regression.

3.4.5 Specifications for Efficacy Endpoint Analyses

3.4.5.1 Data Handling Rules for ORR

Data handling rules for the primary endpoint are as follows:

- Baseline disease evaluations should be performed as close as possible to the study treatment start and never more than 21 days before the beginning of the study treatment.
- An adequate radiographic disease assessment will be defined as one where sufficiently clear radiological images are obtained so that changes in all the target lesions, non-target lesions and pathological lymph nodes may be observed and/or measured, or the presence of new lesions may be determined. A patient will be considered as not evaluable for response at a protocol-specified time point if no imaging/measurement is done or the assessment is not adequate (e.g., only a subset of lesion measurements are made) unless a convincing argument can be made that the contribution of the individual missing lesion(s) would not change the assigned time point response.
- The best overall response is the best response recorded across all time points from the start of the study treatment until the end-of-treatment disease assessment taking into account any requirement for confirmation (i.e., in the determination of best confirmed overall response versus best overall response). Response assessed after disease progression will not be considered in determination of the best overall response. Post-study-treatment radiographic assessments including those after a new anticancer treatment has started will not be considered in determination of the best overall response.
- Time point overall response and the best overall response will be calculated according to Table 3 – 5 in Appendix 2 (RECIST v1.1). A patient will be considered to have a response if the criteria for response have been met at the protocol-specified time points immediately before and after the time point of inevaluable response, e.g., a patient with time point overall responses of PR-NE-PR as a confirmed PR. A single occurrence of SD in the same pattern will also be defined as a confirmed PR. However, a PR followed by

more than one occurrence of either NE or SD will not be considered as confirmed. In such cases subsequent responses can be confirmed.

- If SD is the best time point overall response however the interval from baseline to the radiographic assessment is less than 6 weeks (allowing a minus 3-day time window), the best overall response will depend on the subsequent assessments. A patient who has SD at first radiographic assessment, PD at second and does not meet minimum duration for SD, will have a best response of PD. A patient lost to follow-up after the first radiographic assessment which is SD however the duration of SD is less than 6 weeks (allowing a minus 3-day time window) from the first dose date, would be considered inevaluable.
- All patients will be assigned to one of the following best overall response categories: CR, PR, SD, PD, early death, or unknown (including no measureable disease at baseline, inevaluable post-baseline response because of early discontinuation due to adverse events or other reasons, and no exposure to study drug after randomization). Symptomatic deterioration is not a descriptor of an objective response. All patients whose best response is not CR or PR will be considered as non-responders in the calculation of ORR.
- For the analysis of intracranial CNS ORR and PFS: If a patient progresses due to lesions outside the intracranial CNS and continues on study treatment, this patient will continue to be evaluated as SD, PR, or CR in the intracranial CNS until progression in the intracranial CNS or discontinuation from the study treatment.

3.4.5.2 Determination of Disease Progression

Disease progression will be determined to have occurred based on:

- The appearance of one or more new lesions.
- An increase in the size of target measurable lesions (greater than or equal to 20% of the sum of the longest diameters and at least 5 mm absolute increase per RECIST v1.1).
- A clear, unequivocal increase in non-target disease, an overall level of substantial worsening in non-target disease (whether measurable or non-measurable) of a magnitude, even in the presence of SD or PR in target disease. A simple increase in one or more non-target lesions alone is not considered 'unequivocal progression'.
- In patients with no measurable disease at baseline, disease progression is defined as the appearance of new lesions or substantial worsening in non-target disease.

3.4.5.3 Data handling Rules Specifically for the Primary Analysis of PFS

Data handling rules for the primary PFS analyses will be as follows:

- All PFS events are based on well-documented and verifiable progression events or death.
- Disease progression dates are based on adequate radiographic assessments by the investigator. The date of progression is defined as the date at which disease progression was first evident. For progression based on a new lesion, the progression date is the date of the first observation that the new lesion was detected. If multiple assessments based on the sum of target lesion measurements are done at different times, the progression date is the date of the last observation or radiological assessment of target lesions that shows a predefined increase in the sum of the target lesion measurements. Patients may continue the study treatment beyond documented progressive disease per RECIST v1.1 at the

investigator's discretion. Among these patients, only the date of the first documented evidence of progressive disease will be used to define the PFS.

- If a patient experiences only symptomatic progression or clinical deterioration in the absence of radiologic progression, symptomatic deterioration or clinical progression will not be considered as a disease progression event and therefore is not a PFS event.
- In any efficacy evaluation, PFS for patients still being followed for progression will be censored at the last adequate radiographic progression-free assessment date before the clinical cut-off date.
- The detailed scheme of progression and censoring for the primary analysis of PFS is specified in Table 2.

These rules are defined consistently with FDA Guidance: "FDA Guidance for Industry, Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics" (May 2007).

Table 2 The Scheme of Progression and Censoring for the Primary Analysis of PFS

#Rule	Situation	Date of progression or censoring	Outcome
1	Patient randomized but untreated due to death or any other reason		Censored, PFS=1 day
2	No baseline disease assessment		Censored , PFS=1 day
3	No measurable disease at baseline	Date of new lesion(s) or substantial worsening in non-target disease Date of last adequate progression-free radiographic assessment	Progressed Censored
4	No progression or death	Date of last radiological assessment of measured lesions	Censored
5	Death before first PD assessment	Date of death	Progressed
6	Death between adequate assessment visits	Date of death	Progressed
7	Death after one missed radiographic assessment	Date of death	Progressed
8	Death after two of more missed radiographic assessments	Date of last adequate radiographic assessment	Censored
9	Treatment discontinuation prior to documented disease progression or death	Date of last adequate radiographic assessment	Censored
10	New anticancer treatment started without documented disease progression and lead to treatment discontinuation	Date of last adequate progression-free radiographic assessment prior to initiation of new anticancer treatment	Censored
11	Cancer-related surgery prior to documented disease progression	Date of last adequate progression-free radiographic assessment prior to surgery	Censored
12	Disease progression documented between scheduled visits	Earliest of the following: Date of radiological assessment showing new lesion (if progression is based on new lesion); Date of last radiological assessment of measured lesions (if progression is based on increase in sum of measured lesions)	Progressed
13	Disease progression after 1 missed follow-up disease assessment	Date of progression	Progressed

#Rule	Situation	Date of progression or censoring	Outcome
14	Disease progression after 2 or more missed follow-up disease assessments (death at any time)	Date of last adequate progression-free radiographic assessment	Censored

3.4.5.4 Data handling Rules Specifically for the Sensitivity Analyses of PFS

Sensitivity Analysis I of PFS will be based on disease progression verified by IRC and use the progression and censoring scheme in Table 2, Section 3.4.5.3. In an interim evaluation, for a patient who has progressed per investigator but has not been evaluated by the IRC, PFS will be censored at the last progression-free assessment date as assessed by IRC.

Sensitivity Analysis II of PFS will be performed to assess the impact of missed scheduled disease assessment based on disease progression determined by the investigator. This analysis will use the progression and censoring scheme in Table 2, Section 3.4.5.3 except for Rule #6, #7, #13, and # 14. In this analysis, the death date or the date at which disease progression was first evident will be used as disease progression date no matter how many scheduled disease assessments are missed prior to the death or documented disease progression.

3.4.5.5 Data handling Rules Specifically for the Primary Analysis of Intracranial CNS PFS

In patients with active brain metastases in the intracranial CNS at enrollment, intracranial CNS objective response (PD, SD, PR, CR) will be determined by restricting the RECIST v1.1 criteria to lesions in the intracranial CNS only (target, non-target, and new lesions). A new lesion in the intracranial CNS will be scored as PD in the intracranial CNS. If a patient progresses due to lesions outside the intracranial CNS and continues on study treatment, the patient will continue to be evaluated as SD, PR, or CR in the intracranial CNS until disease progression in the intracranial CNS or the patient discontinues study treatment. Confirmation of CR or PR should be performed 4 weeks (allowing a minus 3-day window) or more after initial response.

The primary analysis of intracranial CNS PFS will be based on radiographic assessments in the intracranial CNS by the IRC. The primary analysis will only include randomized patients with active brain metastases at baseline. In all the analyses, the progression and censoring scheme in Table 2, Section 3.4.5.3 will be used except that the radiographic assessments are restricted to the disease in the CNS.

3.4.5.6 Data Handling Rules Specifically for the Analysis of Duration of Response

The analysis of response duration will be based on disease assessment by the investigator and use the progression and censoring scheme in Table 2, Section 3.4.5.3. The analysis of response duration will only include randomized patients with confirmed CR or PR. An additional analysis of duration of response will be performed using disease assessment performed by the IRC. Censoring in the analysis of duration of response will be the same as for PFS.

3.4.5.7 Data Handling Rules Specifically for the Analysis of OS

All the randomized patients will be included in the primary analysis. For patients who have been randomized and are not treated, the OS will be defined as the time interval from randomization date to death date or date of the last contact if available. In the case that at least 3 patients are randomized and untreated, a sensitivity analysis will be performed for treated patients only.

3.4.5.8 Handling Rules for Response Contemporaneous Scans

Two or more scans may be taken within a few days. The scans may cover multiple anatomical regions and can be of different modalities (e.g. CT scan and MRI). Scans within a 2-day window

will be "merged" lesion by lesion. If a lesion (target, non-target, or new lesion) is on only 1 scan, evaluations of the multiple lesions will be simply combined.

If a single target lesion appears on 2 or more scans, the lesion measurements will be averaged. Any new lesion on any of the multiple scans will be counted as a new lesion. If a non-target lesion is present on multiple scans, the assessment should be the 'worst' of the multiple evaluations, e.g. non-CR/non-PR and absent counts as non-CR/non-PR, progression and non-CR/non-PR counts as progression etc.

3.5 Safety Analysis

Throughout the study, safety assessments will include physical and laboratory examinations, vital signs, and ECGs according to the Schedule of Events. Adverse events will be assessed and categorized by the US National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE), v 4.0 (see protocol Appendix A). Patients will remain on treatment until they meet one or more criteria for withdrawal. Patients will be supplied with study treatment until they discontinue from the study.

All safety analyses will be performed using the treated population, which will include all patients who have received at least one dose of study treatment.

3.5.1 Adverse Events

All adverse events (AEs) starting/worsening on or after the first dose of study treatment and no later than 30 days after the last dose date will be considered as treatment-emergent. Treatment-emergent AEs also include those with partially or completely missing start date since there is not enough evidence that the event started before the first dose of study treatment.

All AEs entered in the clinical database (including treatment-emergent and non-treatment-emergent AEs) will be listed in by-patient listings or available for review in appropriate datasets. The incidence rates of treatment-emergent AEs, as well as the frequency of occurrence of overall toxicity categorized by maximum toxicity grades (severity), will be described. In addition, treatment-emergent AEs will be summarized by causal relationship to study treatment (in the opinion of the investigator) and action taken on study treatment, including dose modifications, interruptions and discontinuation. Serious treatment-emergent AEs, both overall and by causal relationship to study treatment, will also be summarized. Treatment-emergent AEs will also be summarized using cumulative incidence rate by cycles for all treated patients and among treated patients still at risk. Incidence rate for treatment-emergent AEs and AE groups in special categories, e.g., pulmonary adverse events overall and within first 7 days, will be summarized and also be adjusted with patients' entire exposure to study treatment and exposure through the AE initial onset. AEs in special categories will also be summarized in terms of number of occurrences, by time to initial onset and duration of events. The special categories to be analyzed will include at minimum those listed in Appendix 4. All deaths occurring during a period of exposure to AP26113 or within a period of up to 30 days following discontinuation from AP26113, and all deaths occurring later but resulting from treatment related adverse event(s) will be summarized by AEs leading to death and causal relationship to study treatment.

The incidence rates and cumulative incidence rate of treatment-emergent AEs will also be summarized in subgroups based on selected demographic and baseline characteristics, randomization stratification factors, and the mutation type in ALK and other selected

biomarkers. The correlation of ALK and other biomarkers with the incidence rates of selected treatment-emergent AEs will be evaluated using logistic regression.

3.5.1.1 Imputation Rule for Missing Causal Relationship of Treatment-Emergent Adverse Events

Missing causal relationship to the study treatment will be imputed as the least possible related for treatment-emergent AEs.

3.5.1.2 Imputation Rules for Missing Onset Date and Resolution Date of Adverse Events

In general the imputation will be conservative such that onset dates will be imputed to be as early as possible and resolution dates will be imputed to be as late as possible. Resolution date will be imputed first and then used to impute onset date.

Imputation for Resolution Date:

- If day is missing but month and year are non-missing (UU-MMM-YYYY), impute as the earliest of:
  - Last day of the month (28, 29, 30 or 31 depending on in which month the adverse event resolved)
  - Data cutoff date
  - Death date
- If day and month are missing (UU-UUU-YYYY), impute as the earliest of:
  - December 31 (31-DEC-YYYY)
  - Data cutoff date
  - Death date
- If date is completely missing (e.g., AE is ongoing), impute as earliest of:
  - Data cutoff date
  - Treatment discontinuation date + 30 days
  - Death date

Imputation for Onset Date:

- If day is missing but month and year are non-missing (UU-MMM-YYYY), impute as follows:
  - If year and month are the same as year and month of first dose date:
    - If resolution date (or imputed resolution date) is on or after first dose date, impute as first dose date
    - If resolution date (or imputed resolution date) is prior to first dose date, impute as latest of first day of the month or informed consent date
  - If year is the same as year of first dose date and month is **after** month of first dose date, impute as first date of month
  - If year is the same as year of first dose date and month is **before** month of first dose date, impute as latest of first day of the month or informed consent date
  - If year is **after** year of first dose date, impute as first day of month
  - If year is **before** year of first dose date, impute as latest of first day of the month or informed consent date
- If day and month are missing and year is non-missing (UU-UUU-YYYY), impute as follows:

- If year is the same as year of first dose date:
  - If resolution date (or imputed resolution date) is on or after first dose date, impute as first dose date
  - If resolution date (or imputed resolution date) is prior to first dose date, impute as latest of first day of the month or informed consent date
- If year is **after** year of first dose date, impute as January 1 (01-JAN-YYYY)
- If year is **before** year of first dose date, impute as latest of first day of the year (01-JAN-YYYY) or informed consent date

- If date is completely missing:
  - If resolution date (or imputed resolution date) is on or after first dose date, impute as first dose date
  - If resolution date (or imputed resolution date) is prior to first dose date, impute as informed consent date.

3.5.1.3 Derivation of Treatment-Emergent Adverse Event Flag

Sites are instructed to "split" AEs that are ongoing as of first dose date into two records: one that ends as of first dose date and one that begins on first dose date. Therefore in some cases, what may at first appear as an AE that begins on first dose date is really a continuation of an AE that began prior to first dose date. If all of the following criteria are true for any two AE records then the two records should be considered as one event:

- The preferred term of the first AE = the preferred term of the second AE.
- The start date of the first AE is before the date of first dose.
- The end date of the first AE is one day before the date of first dose or equal to the date of first dose.
- The start date of the second AE = the date of first dose.
- The severity of the second AE <= the severity of the first AE.
- The relatedness of both the first and the second AEs is "Not Related".

After determining which records should be treated as a single event, the treatment-emergent flag can be determined as follows:

- All AEs with an onset date on or after the first dose date and no later than 30 days after the last dose date.
- In the case of missing onset date, impute using the rules defined in Section 3.5.1.2

3.5.1.4 Calculation Specifications for Treatment-Emergent Adverse Events

Time to onset is defined as the time interval from the first dose date until the onset date of a treatment-emergent AE. In case of a missing onset date, a date imputed with the rules in Section 3.5.1.2 will be used. Time to onset will be computed with the following formula: time to onset= AE onset date – first dose date + 1.

Time to initial onset of an individual treatment-emergent AE is defined as the time interval from the first dose date until the earliest onset date of the treatment-emergent AEs of the same nature. In case of a missing onset date, a date imputed with the rules in Section 3.5.1.2 will be used. Time to initial onset will be computed with the following formula: time to initial onset = earliest AE onset date – first dose date + 1.

Age at onset will be computed by the following formula and rounding down to the nearest integer: age at onset = floor ((AE start date – birth date + 1)/365.25).

Dose at onset will be defined as the dose received by patients on AE onset date, regardless of treatment arm. Dose at onset will not be defined if the occurrence of a treatment-emergent AE is post treatment discontinuation.

Dose by onset will be defined as the last non-zero dose received by AE onset date.

3.5.1.5 Specifications for the Analyses Outputs

Adverse events will be coded with MedDRA dictionary. The version of the dictionary will be specified in the output listings or tables. Some preferred terms coding manifestations of similar medical conditions will be re-coded and re-grouped based on the synonym infrastructure determined through sponsor medical review (An example of such an infrastructure is provided in Appendix 3). In the summary tables, MedDRA preferred terms will be sorted in a descending order of incidence rate first then alphabetically in case of tied incidence rates. In the tables of treatment emergent AEs grouped by MedDRA System Organ Class (SOC), SOCs will be put in the internationally agreed order.

3.5.2 Laboratory and Vital Signs Data

Laboratory and vital signs data in standard units will be summarized using summary statistics and graphically at baseline and at each cycle for which adequate data are available. Laboratory and blood pressures data will also be graded according to the NCI CTCAE, v.4.0 when applicable. Change from baseline to the worst on-study result will be summarized using the changes from baseline values in standard units. Change from baseline to the worst on-study result will also be summarized by shift in CTCAE grade for selected laboratories and blood pressures. Changes in Testosterone and Insulin will be summarized by cross-tabulations of baseline (low, normal, high) and the highest/lowest value on study (low, normal, high).

Laboratory measurements of ALT, AST, Alkaline phosphate (ALP), and total bilirubin (TBL) will be evaluated for the potential risk of drug-induced liver failure.

Possible Hy's Law cases (Zimmerman 1999) are defined as patients with ALT or AST > 3 × ULN, with ALP < 2 × ULN and total bilirubin (TBL) ≥ 2 × ULN with no other etiology to explain these liver-function test results. Changes in ALT, AST, ALP, and TBL over time will also be graphically displayed for possible Hy's law patients.

3.5.3 QTcF Analysis

Descriptive statistics of maximum QTcF and change from baseline will be calculated following the ICH-E14 guidelines: the proportion of treated patients with at least one on-drug QTcF value > 450 ms, 480 ms, and 500 ms; proportion of treated patients with a maximum change in QTcF from baseline > 30 ms and > 60 ms. The mean change in QTcF from baseline to maximum on study value will be calculated along with 95% confidence intervals.

3.5.4 Extent of Exposure

Exposure to AP26113 will be summarized using the following measures:

- Time (days) on study treatment

- Total cumulative dose of AP26113 administered
- Dose intensity (mg/day)
- Relative dose intensity (%)
- Total person years
- Dose interruption of at least 3 days
    - Number of patients with at least one occurrence
    - Number of patients who returned to dosing after interruption
    - Number of patients who returned to target dose after interruption
    - Total duration (days) of time off study drug prior to treatment discontinuation
- Dose reduction of at least 3 days
    - Number of patients with at least one occurrence
    - Number of patients who returned to dosing after interruption
    - Number of patients who returned to target dose after interruption
    - Total duration (days) of time off study drug prior to treatment discontinuation

Time on treatment will be defined as the time interval from the first dose date to the last dosing date and computed with the following formula: Time (days) on treatment= last non-zero dose date – first dose date + 1

Dose intensity will be calculated with the following formula: Dose intensity = Total cumulative dose/ Time (days) on study treatment. Relative dose intensity will be defined as the proportion of the planned dose received by patients. In treatment arm A, daily planned dose will be 90 mg and change to the escalated dose since the day 1 of escalation. In treatment arm B, daily planned dose will be 90 mg in the 7-day lead-in period and 180 mg from day 8 onward. Relative dose intensity will be calculated as follows: Relative dose intensity = Total cumulative dose administered / Total dose planned x 100%.

Total person years for a treated patient will be calculated using the following formula:

Total person years = Time (days) on study treatment / 365.25. The total person years in an analysis population will be the sum of the total person years of all the patients in this population.

Dose modifications will be summarized by dose interruption and dose reduction. A patient will be identified as having dose interruption if this patient had no exposure to study drug for at least 3 consecutive days. A patient will be identified as having dose reduction if this patient had a period of reduced dosage of at least 3 consecutive days, as long as the dose received was less than the target dose but greater than 0 mg on some of the days in this period. Periods of time in which a patient alternates between reduced dosing and dose interruptions will be handled in the following manner:

- The entire period of time between the last receipt of the target dose and either resumption at the target dose or discontinuation of treatment will be considered as a single dose reduction period.
- Any period of 3 or more days with no receipt of study drug within that dose reduction period will also be treated as a dose interruption.

3.6 Quality of Life Measurement

Patient-reported symptoms and HRQoL will be collected by administering the EORTC QLQ C30 (v.3.0, Aaronson et al., 1993) questionnaire at pre-specified scheduled visits, and at the visit

30 days after the last dose of AP26113. The PRO questionnaire will be administered to patients when they arrive for their scheduled visits, prior to any clinical measurements, assessments, evaluations, or procedures being performed.

The EORTC QLQ-C30 will be scored for 5 functional scales (physical, role, cognitive, emotional, and social functioning); 3 symptom scales (fatigue, pain, and nausea/vomiting); and a global health status/QoL scale. Six single item scales also are included (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).

3.6.1 Analyses of Quality of Life Measurements

For HRQoL measures, raw scores for multi-item scales will be calculated by averaging items within scales first. Raw scores will be summarized by time point with descriptive statistics for each scale. Raw scores for multi-item scales and single-item measures will be linearly transformed to obtain the score ranging from 0 to 100 according to EORTC QLQ-C30 (V3) Scoring Manual (Fayers et al., 2001). The global health status / QoL scale based upon Q29 and Q30 will be used as the overall summary measure.

The HRQoL scores including the overall summary measure will be summarized at baseline and by time point in evaluable patients overall and by treatment group. The changes from baseline over time will be summarized with descriptive statistics and explored using mixed effects models.

Information collected in the EORTC QLQ-C30 will be analyzed on its own, and findings such as dyspnea collected as part of the EORTC QLQ-C30 will not be considered as AEs.

3.6.2 Handling Rule of Missing Quality of Life Measurements

Handling rules for missing QoL measurement are as follows:

- Missing items will be imputed as the average of the items which are present for a multi-item scale if at least half of the items from the scale have been answered.
- A missing single-item measure will not be imputed.
- Missing forms will not be imputed.
- Patients with missing baseline scores will be excluded from the analysis for a scale when the change from baseline is analyzed or the baseline score is used as a covariate.

3.7 Pharmacokinetic Analysis

All patients must provide a blood sample (approximately 3 mL per sample) for analysis of steady-state AP26113 plasma concentration. Samples will be collected on Day 1 of Cycles 2, 3, 4, and 5. Cycle 2 requires sampling at pre-dose, and at 1 hour (±10 minutes), 4 hours (±15 minutes) and 6-8 hours (±15 minutes) post-dose. During Cycles 3, 4, and 5, samples will be collected at two time points from each patient (a pre-dose sample and a second sample anytime between 1 to 8 hours post-dose). The AP26113-13-201 study samples will be analyzed with population PK modeling in combination with data from the "AP26113-11-101" study. PK data will also be used in an exposure-response analysis for safety and efficacy. The exposure-response relationship between the simulated concentration (AUC and C-24hr) and ORR and PFS will be analyzed with stepwise logistic regression and stepwise Cox proportional regression models with adjusting potential confounding factors. The exposure-response analysis for

selected adverse events and/or laboratory assessments will be analyzed in a similar manner. PK analyses will be detailed in a statistical analysis plan specifically developed for this purpose.

4 REFERENCES

Aaronson NK, Ahmedzai S, Bergman, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365-376.

Kaplan EL and Meier P. Nonparametric Estimation from Incomplete observations. J Am Statistical Assoc. 1958;53:457-481.

Fayers PM, Aaronson NK, Bjordal K, et al. EORTC Scoring Manual. 3rd edition. Brussels: European Organisation for Research and Treatment of Cancer; 2001.

"FDA Guidance for Industry, Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics" (May 2007) available at http://www.fda.gov/downloads/Drugs/.../Guidances/ucm071590.pdf.

5 APPENDICES

The final SAP will include the following Appendices

Appendix 1 Schedule of Events

	Screening Period	Treatment through 30 Days After Last Dose								Follow-up Period
Assessment	**Screening[1]**	**Cycle 1 (28 Days)**			**Every 4 weeks (Day 1 of each cycle from Cycle 2 onwards)**	**Every 8 weeks (from Day 1 of Cycle 3 through Cycle 15; every 3 cycles thereafter)**	**Every 12 weeks (from Day 1 of 3rd Cycle from Cycle 4 onwards)**	**End-of-Treatment[27]**	**30 Days After Last Dose[28]**	**Follow-up[29]**
Day (D)	**D-14 to D0**	**D1**	**D8**	**D15**						
Informed Consent[2]	X									
Randomization[3]	X									
Demographics[4]	X									
Medical/Surgical History[5]	X									
Diagnosis and Cancer History[6]	X									
Prior Cancer Therapy[7]	X									
ALK Mutation Status[8]	X									
Physical Examination[9]	X	X	X*	X*	X			X	X	
Vital Signs[10]	X	X	X	X	X			X	X	
ECOG Performance Status[11]	X	X			X			X	X	
Hematology[12]	X	X		X	X			X	X	
Chemistry (fasting, if possible)[13]	X	X		X	X			X	X	
Urinalysis[14]	X	X			X			X	X	
Insulin (fasting, if possible)[15]	X	X		X	X			X	X	
Testosterone Level (males only)[16]	X	X			X			X	X	

	Screening Period	Treatment through 30 Days After Last Dose								Follow-up Period
Assessment	**Screening[1]**	**Cycle 1 (28 Days)**			**Every 4 weeks (Day 1 of each cycle from Cycle 2 onwards)**	**Every 8 weeks (from Day 1 of Cycle 3 through Cycle 15; every 3 cycles thereafter)**	**Every 12 weeks (from Day 1 of 3rd Cycle from Cycle 4 onwards)**	**End-of-Treatment[27]**	**30 Days After Last Dose[28]**	**Follow-up[29]**
Day (D)	**D-14 to D0**	**D1**	**D8**	**D15**						
Prothrombin Time (PT)/Partial Thromboplastin Time (PTT)[17]	X	X		X	X			X	X	
Electrocardiogram (ECG)[18]	X	X			X			X	X	
Adverse Events[19]	Throughout Study									
Concomitant Treatments[20]	Throughout Study									
Pregnancy Test[21]	X	X					X	X		
Disease Assessment CT/MRI scans[22]	X					X		X		
Brain MRI[22]	X					X				
Tissue for ALK FISH Testing[23]	X									
Post-Crizotinib Tissue for Molecular Genetics[23]	X									
End-of-Treatment Tissue Sample for Molecular Genetics[23]								X		
CCI [redacted]										
Plasma Sample for Steady-state AP26113 Concentration[25]					X					

	Screening Period	Treatment through 30 Days After Last Dose								Follow-up Period
Assessment	**Screening[1]**	**Cycle 1 (28 Days)**			**Every 4 weeks (Day 1 of each cycle from Cycle 2 onwards)**	**Every 8 weeks (from Day 1 of Cycle 3 through Cycle 15; every 3 cycles thereafter)**	**Every 12 weeks (from Day 1 of 3rd Cycle from Cycle 4 onwards)**	**End-of-Treatment[27]**	**30 Days After Last Dose[28]**	**Follow-up[29]**
Day (D)	**D-14 to D0**	**D1**	**D8**	**D15**						
Patient-Reported Outcomes Assessment[26]	X	X			X			X	X	
Subsequent Anticancer Therapy/Survival[27]										X

*Assessment for early pulmonary symptoms must be performed during the visit on Day 8 and Day 15.

Appendix 2 Notes from RECIST v1.1

Appendix 2.1 Measureable disease vs. Non-measureable disease

Measurable disease is defined by the presence of at least one lesion with a longest diameter > = 10 mm or a pathological lymph node with a short axis of > = 15 mm. Lesions with prior local treatment are usually not considered measurable unless there has been demonstrated progression in the lesion.

Non-measurable disease is defined as small lesions (longest diameter <10mm or pathological lymph nodes with > = 10 to < 15 mm short axis) or lesions truly non-measurable by reproducible imaging techniques.

Appendix 2.2 Target lesions vs. Non-target lesions vs. New lesions

1. Target lesions
    a. In the systemic assessment of disease, a maximum of five total (and from the five, up to two per organ, maximum) lesions or pathological lymph nodes will be identified at baseline for follow up to assess tumor burden in target lesions for response determination (see Section 3.4.4 for discussion of intracranial CNS assessment) . The sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all target lesions will be calculated as baseline or on-study sum diameters.
    b. Measurement of target lesions:
        - "Too small to measure" - In case that target lesions become very faint on CT scan and are reported as 'too small to measure': if it is the opinion of the radiologist that a lesion has likely disappeared, the measurement should be recorded as 0mm; if a lesion or lymph node is believed to be present and is faintly seen but too small to measure, a default value of 5mm should be assigned. This default value is derived from the 5mm CT slice thickness (but should not be changed with varying CT slice thickness). If the radiologist is able to provide an actual measure, that should be recorded, even if it is below 5mm.
        - Lesions that split or coalesce on treatment - When non-nodal lesions 'fragment', the longest diameters of the fragmented portions should be added together to calculate the target lesion sum. Similarly, as lesions coalesce, a plane between them may be maintained that would aid in obtaining maximal diameter measurements of each individual lesion. If the lesions have truly coalesced such that they are no longer separable, the vector of the longest diameter in this instance should be the maximal longest diameter for the 'coalesced lesion'.

c. Response criteria for target lesions:
   - Complete Response (CR): Disappearance of all target lesions. Any pathological lymph nodes (whether target or non-target) must have reduction in short axis to < 10 mm.
   - Partial Response (PR): At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters.
   - Progressive Disease (PD): At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. (Note: the unequivocal appearance of one or more new lesions is also considered progression).
   - Stable Disease (SD): Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters while on study.

2. Non-target lesions
   a. All other lesions or pathological lymph nodes (or sites of disease) after the target lesions are chosen will be identified as non-target lesions and should also be recorded at baseline.
   b. Measurements are not required and these lesions will be assessed only qualitatively as 'present', 'absent', or in rare cases 'unequivocal progression' (more details to follow). In addition, it is possible to record multiple non-target lesions involving the same organ as a single item on the case record form (e.g. 'multiple enlarged pelvic lymph nodes' or 'multiple liver metastases').
   c. Response criteria for non-target lesions
      - Complete Response (CR): Disappearance of all non-target lesions and normalization of tumor marker level. All lymph nodes must be non-pathological in size (<10mm short axis).
      - Non-CR/Non-PD: Persistence of one or more non-target lesion(s) and/or maintenance of tumormarker level above the normal limits.
      - Progressive Disease (PD): Unequivocal progression of existing non-target lesions. (Note: the appearance of one or more new lesions is also considered progression).
   d. Assessment of progression of non-target lesions
      - When the patient has measurable disease at baseline, the designation of overall progression solely on the basis of change in non-target disease in the face of SD or PR of target disease is extremely rare. To achieve 'unequivocal progression' on the basis of the non-target disease, there must be an overall level of substantial worsening in non-target disease such that, even in presence of SD or PR in target disease, the overall tumor burden has increased sufficiently and requires discontinuation of treatment.

- When the patient has only non-measurable disease at baseline, there is no measurable disease assessment to factor into the interpretation of an increase in non-measurable disease burden. An unequivocal progression is to consider if the increase in overall disease burden based on the change in non-measurable disease is substantial and comparable in magnitude to the increase that would be required to declare PD form measurable disease: e.g., an increase in tumor burden representing an additional 73% increase in 'volume' (which is equivalent to a 20% increase diameter in a measurable lesion).

3. New lesions

   A lesion identified on a follow-up study in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression. If a new lesion is equivocal, for example because of its small size, continued therapy and follow-up evaluation will clarify if it represents truly new disease. If repeat scans confirm there is definitely a new lesion, then progression should be declared using the date of the initial scan.

Appendix 2.3 Time point overall response status and best overall response status calculation

Table 3 Summary of the Overall Response Status Calculation at each Time Point for Patients Who have Target +/- Non-target Disease at Baseline

Target Lesions	Non-target Lesions	New Lesions	Overall Response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
Not all evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = inevaluable.

Table 4 Summary of the Overall Response Status Calculation at each Time Point for Patients Who Have Non-measurable (Therefore Non-target) Disease only at Baseline

Non-target Lesions	New Lesions	Overall Response
CR	No	CR
Non-CR/non-PD	No	Non-CR/non-PD[a]
Not all evaluated	No	NE
Unequivocal PD	Yes or No	PD
Any	Yes	PD

CR = complete response, PD = progressive disease, and NE = inevaluable.

[a] 'Non-CR/non-PD' is preferred over 'stable disease' for non-target disease since SD is increasingly used as endpoint for assessment of efficacy in some trials so to assign this category when no lesions can be measured is not advised.

Table 5 Summary of the Best Overall Response Status Calculation when Confirmation of CR and PR is Required

Overall response First time point	Overall response Subsequent time point	Overall Response
CR	CR	CR
CR	PR	SD, PD or PR[a]
CR	SD	SD provided minimum criteria for SD duration met, otherwise PD
CR	PD	SD provided minimum criteria for SD duration met, otherwise PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = inevaluable.

[a] If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the patient had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

Appendix 3 Adverse Events Synonyms*

MedDRA Dictionary Preferred Term	MedDRA Dictionary System Organ Class	Recoded Preferred Term	Recoded System Organ Class
Thrombocytopaenia	Blood and Lymphatic System Disorders	platelet count decreased	Investigations
Thrombocytopenia	Blood and Lymphatic System Disorders	platelet count decreased	Investigations
Thrombocytosis	Blood and Lymphatic System Disorders	platelet count increased	Investigations
Thrombocythaemia	Blood and Lymphatic System Disorders	platelet count increased	Investigations
Thrombocythemia	Blood and Lymphatic System Disorders	platelet count increased	Investigations
Anemia	Blood and Lymphatic System Disorders	Anaemia	Blood and Lymphatic System Disorders
Haemoglobin decreased	Investigations	Anaemia	Blood and Lymphatic System Disorders
Hemoglobin decreased	Investigations	Anaemia	Blood and Lymphatic System Disorders
Haematocrit decreased	Investigations	Anaemia	Blood and Lymphatic System Disorders
Erythrocytosis	Blood and Lymphatic System Disorders	Hemoglobin increased	Investigations
Erythrocythemia	Blood and Lymphatic System Disorders	Hemoglobin increased	Investigations
Leukopenia	Blood and Lymphatic System Disorders	white blood cell count decreased	Investigations
white blood cell count increased	Investigations	Leukocytosis	Blood and Lymphatic System Disorders
Neutropenia	Blood and Lymphatic System Disorders	Neutrophil count decreased	Investigations
Lymphopaenia	Blood and Lymphatic System Disorders	Lymphocyte count decreased	Investigations
Lymphopenia	Blood and Lymphatic System Disorders	Lymphocyte count decreased	Investigations
eosinophil count increased	Investigations	Eosinophilia	Blood and Lymphatic System Disorders
basophil count increased	Investigations	Basophilia	Blood and Lymphatic System Disorders
blood glucose decreased	Investigations	Hypoglycaemia	Metabolism and Nutrition Disorders
Hypoglycemia	Metabolism and Nutrition Disorders	Hypoglycaemia	Metabolism and Nutrition Disorders
MedDRA Dictionary Preferred Term	MedDRA Dictionary System Organ Class	Recoded Preferred Term	Recoded System Organ Class
blood glucose increased	Investigations	Hyperglycaemia	Metabolism and Nutrition Disorders

MedDRA Dictionary Preferred Term	MedDRA Dictionary System Organ Class	Recoded Preferred Term	Recoded System Organ Class
Hyperglycemia	Metabolism and Nutrition Disorders	Hyperglycaemia	Metabolism and Nutrition Disorders
blood sodium decreased	Investigations	Hyponatraemia	Metabolism and Nutrition Disorders
Hyponatremia	Metabolism and Nutrition Disorders	Hyponatraemia	Metabolism and Nutrition Disorders
blood sodium increased	Investigations	Hypernatraemia	Metabolism and Nutrition Disorders
Hypernatremia	Metabolism and Nutrition Disorders	Hypernatraemia	Metabolism and Nutrition Disorders
blood potassium decreased	Investigations	Hypokalaemia	Metabolism and Nutrition Disorders
Hypokalemia	Metabolism and Nutrition Disorders	Hypokalaemia	Metabolism and Nutrition Disorders
blood potassium increased	Investigations	Hyperkalaemia	Metabolism and Nutrition Disorders
Hyperkalemia	Metabolism and Nutrition Disorders	Hyperkalaemia	Metabolism and Nutrition Disorders
blood calcium decreased	Investigations	Hypocalcaemia	Metabolism and Nutrition Disorders
Hypocalcemia	Metabolism and Nutrition Disorders	Hypocalcaemia	Metabolism and Nutrition Disorders
Hypercalcemia	Metabolism and Nutrition Disorders	Hypercalcaemia	Metabolism and Nutrition Disorders
blood calcium increased	Metabolism and Nutrition Disorders	Hypercalcaemia	Metabolism and Nutrition Disorders
blood magnesium decreased	Investigations	Hypomagnesaemia	Metabolism and Nutrition Disorders
Hypomagnesemia	Metabolism and Nutrition Disorders	Hypomagnesaemia	Metabolism and Nutrition Disorders
blood magnesium increased	Investigations	Hypermagnesaemia	Metabolism and Nutrition Disorders
Hypermagnesemia	Metabolism and Nutrition Disorders	Hypermagnesaemia	Metabolism and Nutrition Disorders
Blood Phosphorus decreased	Investigations	Hypophosphataemia	Metabolism and Nutrition Disorders
blood phosphate decreased	Investigations	Hypophosphataemia	Metabolism and Nutrition Disorders
MedDRA Dictionary Preferred Term	MedDRA Dictionary System Organ Class	Recoded Preferred Term	Recoded System Organ Class
Hypophosphatemia	Metabolism and Nutrition Disorders	Hypophosphataemia	Metabolism and Nutrition Disorders
blood phosphorous increased	Investigations	Hyperphosphataemia	Metabolism and Nutrition Disorders
Hyperphosphatemia	Metabolism and Nutrition Disorders	Hyperphosphataemia	Metabolism and Nutrition Disorders
blood triglycerides increased	Investigations	Hypertriglyceridaemia	Metabolism and Nutrition Disorders

MedDRA Dictionary Preferred Term	MedDRA Dictionary System Organ Class	Recoded Preferred Term	Recoded System Organ Class
Hypertriglyceridemia	Metabolism and Nutrition Disorders	Hypertriglyceridaemia	Metabolism and Nutrition Disorders
Hypercholesterolemia	Metabolism and Nutrition Disorders	Blood cholesterol increased	Investigations
Hypercholesterolaemia	Metabolism and Nutrition Disorders	Blood cholesterol increased	Investigations
blood albumin decreased	Investigations	Hypoalbuminaemia	Metabolism and Nutrition Disorders
Hypoalbuminemia	Metabolism and Nutrition Disorders	Hypoalbuminaemia	Metabolism and Nutrition Disorders
Hyperbilirubinaemia	Hepatobiliary Disorders	blood bilirubin increased	Investigations
Hyperbilirubinemia	Hepatobiliary Disorders	blood bilirubin increased	Investigations
Blood Lipase increased	Investigations	Lipase Increased	Investigations
Lipase	Investigations	Lipase Increased	Investigations
Blood Amylase	Investigations	Blood Amylase Increased	Investigations
Pancreatitis Acute	Gastrointestinal disorders	Pancreatitis	Gastrointestinal disorders
Rash erythematous	Skin and Subcutaneous Tissue Disorders	Rash	Skin and Subcutaneous Tissue Disorders
Rash Macular	Skin and Subcutaneous Tissue Disorders	Rash	Skin and Subcutaneous Tissue Disorders
Rash papular	Skin and Subcutaneous Tissue Disorders	Rash	Skin and Subcutaneous Tissue Disorders
rash maculo-papular	Skin and Subcutaneous Tissue Disorders	Rash	Skin and Subcutaneous Tissue Disorders
Troponin I increased	Investigations	Troponin increased	Investigations
painful defaecation	Gastrointestinal Disorders	proctalgia	Gastrointestinal Disorders
blood pressure increased	Investigations	hypertension	Vascular disorders

MedDRA Dictionary Preferred Term	MedDRA Dictionary System Organ Class	Recoded Preferred Term	Recoded System Organ Class
blood pressure decreased	Investigations	hypotension	Vascular disorders
escherichia bacteraemia	Infections and Infestations	bacteraemia	Infections and Infestations
staphylococcal bacteraemia	Infections and Infestations	bacteraemia	Infections and Infestations
abdominal pain upper	Gastrointestinal Disorders	abdominal pain	Gastrointestinal Disorders
abdominal pain lower	Gastrointestinal Disorders	abdominal pain	Gastrointestinal Disorders
stomach discomfort	Gastrointestinal Disorders	abdominal discomfort	Gastrointestinal Disorders

*This list will be updated as new events occur and/or MedDRA versions change over the course of the study.

Appendix 4 Adverse Events in Special Categories

Selection of AEs for special categories will be performed through the use of Standardised MedDRA Queries (SMQs), Modified MedDRA Queries (MMQs), or Customized MedDRA Queries (SMQs). The adverse events in special categories will be listed and updated as new events occur and/or MedDRA versions change over the course of the study. The final versions of these query sets will be provided as an appendix to the final CSR.

The following special categories of adverse events will be analyzed:

- Pulmonary: early onset pulmonary events and later onset pneumonitis-like events
- Hypertension events
- Bradycardia events
- Gastrointestinal events
- Hepatic events
- Pancreatic events
- Insulin level elevation/hyperglycemia events
- Skin and soft tissue SOC events