PROTOCOL A4061068

A PHASE 1B, OPEN LABEL, DOSE ESCALATION STUDY TO EVALUATE SAFETY, PHARMACOKINETICS AND PHARMACODYNAMICS OF AXITINIB (AG-013736) IN COMBINATION WITH CRIZOTINIB (PF-02341066) IN PATIENTS WITH ADVANCED SOLID TUMORS

Statistical Analysis Plan

(SAP)

Version: 2.1

Author: PPD

PPD

Date: September 3, 2015

[redacted]

[redacted]

TABLE OF CONTENTS

1. AMENDMENTS FROM PREVIOUS VERSION(S)

- Added analyses of safety data (adverse events and laboratory tests) which includes the period from the start of treatment with the axitinib + crizotinib combination (C1D1) for Dose Escalation cohorts and Dose Expansion cohort 1.

- Clarified language regarding analysis sets for efficacy endpoints.

- Clarified language for analysis of biomarker endpoints.

2. INTRODUCTION

This document describes the planned statistical analyses for Protocol A4061068. This analysis plan is meant to supplement the study protocol. In this document, any text taken directly from the protocol is *italicized*.

Any deviations from this main analysis plan will be described in the Clinical Study Report.

2.1. Study Design

This is a Phase 1b, open-label, multi-center, multiple-dose, safety, pharmacokinetics (PK) and pharmacodynamic study of axitinib in combination with crizotinib in adult patients with advanced solid tumors. This clinical study will be composed of a Dose Escalation Phase and a Dose Expansion Phase. The Dose Escalation Phase will estimate the maximum tolerated dose (MTD) in dose escalation cohorts in patients with advanced solid tumors, using the modified toxicity probability interval (mTPI) method.

The Dose Escalation Phase will lead to the identification of an Expansion Test Dose for axitinib in combination with crizotinib in patients with solid tumors. The Expansion Test Dose will be either the MTD or the Maximum Feasible Dose (MFD), ie., the highest tested dose that is declared safe and tolerable by the Investigators and Sponsor. Once the Expansion Test Dose is identified, the Dose Expansion Phase will be opened and axitinib in combination with crizotinib will be tested in patients with advanced renal cell carcinoma (RCC).

The Dose Expansion Phase is comprised of two patient populations, both with histologically or cytologically confirmed advanced RCC with a component of clear cell subtype, with primary tumor resected and:

- *Cohort 1: No prior systemic therapy directed at advanced RCC.*

- *Cohort 2: At least one but no more than two prior systemic treatment regimens directed at advanced RCC, with the last prior therapy being a single agent VEGF-pathway inhibitor or bevacizumab plus IFN-α, and whose disease has acquired resistance to this treatment. Resistance is defined as disease progression as per response evaluation criteria in solid tumors (RECIST) version 1.1 following either an initial objective response (complete or partial), or stable disease for at least 6 months.*

To understand the effect of crizotinib on axitinib PK, a 7-day lead-in period of single-agent axitinib directly preceding the administration of the crizotinib and axitinib combination will be included prior to Cycle 1 in the Dose Escalation Phase of the study. Axitinib is not expected

to affect crizotinib exposure so there will be no PK study with a lead-in period of single-agent crizotinib dosing for this evaluation. Although not part of the analysis plan of this study, the effect of axitinib on crizotinib PK will be assessed by comparing the crizotinib exposures observed in this study in combination with axitinib with historical data from prior studies that involved administration of crizotinib alone. The Dose Expansion Phase Cohort 1 will be used to further study crizotinib and axitinib PK interactions with PK assessments (from at least 8 evaluable subjects) similar to those conducted in the Dose Escalation Phase, No lead-in period will be included in the Dose Expansion Phase Cohort 2 since no PK collections will be conducted for this cohort.

Anti-tumor activity will be evaluated every 8 weeks after Cycle 1 Day 1 using RECIST version 1.1. Given the safety profile of crizotinib, in addition to standard safety tests, ophthalmology examinations will be carried out to assess any vision changes. Based on the safety profile of axitinib, blood pressure will be monitored throughout the treatment period, as well as thyroid function.

Electrocardiogram (ECG) measurements will be taken throughout the treatment period in all patients and in conjunction with PK sampling in patients from the Dose Escalation Phase and the Dose Expansion Phase Cohort 1. Archived tumor tissues will be collected for all patients. De novo tumor biopsy will be collected for the patients in the Dose Expansion Phase Cohort 2. For all patients a second biopsy might be provided on a voluntary basis at the time of disease progression. Biomarker studies on tumor tissue and blood will be carried out to help understand the mechanism of action of the axitinib plus crizotinib combination, as well as potential mechanisms of resistance. Such results may help in the future development of this combination. These analyses may also result in the identification of potential biomarkers of response to the axitinib plus crizotinib combination, ultimately leading to development of a patient selection strategy for further clinical investigation. As such collection and analysis of the archival tumor tissue as well as de novo tumor biopsies at baseline and at time of progression will be paramount to generate such knowledge.

2.2. Study Objectives

2.2.1. Primary Objective

- *To assess the safety and tolerability of axitinib in combination with crizotinib in patients with solid tumors and advanced RCC in order to estimate the MTD (or MFD) and select the recommended Phase 2 dose (RP2D).*

2.2.2. Secondary Objectives

- *To evaluate the overall safety profile;*

- *To characterize the pharmacokinetics (PK) of axitinib and crizotinib when administered in combination and to assess the effect of crizotinib on the PK of axitinib (Dose Escalation Phase and the 8 PK-evaluable subjects in Expansion Phase Cohort 1 only);*

- *To characterize the effects of axitinib in combination with crizotinib on QTc;*

- *To document the anti-tumor activity of axitinib in combination with crizotinib in advanced RCC patients;*

- *To explore the pharmacodynamic effect of axitinib in combination with crizotinib in blood;*

- *To characterize the alterations and/or expression profiles of genes, proteins, and RNAs relevant to angiogenesis (eg, Ang-2), drug targets (eg, c-MET) and sensitivity and/or resistance (eg, PBRM1) to axitinib in combination with crizotinib in tumor and/or blood.*

3. INTERIM ANALYSES, FINAL ANALYSES, AND UNBLINDING

This is an open label, single-arm trial for which no formal interim analysis is planned. The final analysis will be performed after the last subject last visit; however, earlier analyses of the data may be performed for publication and regulatory reporting purposes.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The emphasis of the final analyses will be on estimation of key summary statistics rather than hypothesis testing.

4.2. Statistical Decision Rules

Up-and-Down Matrix Design with the mTPI Method

The dose escalation/de-escalation rules will follow the mTPI method (see Section 8 for a detailed description). Briefly, the mTPI method relies upon a statistical probability algorithm, calculated using data from all patients treated at the same dose level (and not simply those in the current cohort) to determine whether future cohorts should involve dose escalation, no change in dose, or dose de-escalation.

Maximum Tolerated Dose Definition
The MTD estimate is the highest dose of axitinib and crizotinib associated with the occurrence of DLTs in <33% of patients.

Dose Expansion Phase Cohorts

Once the MTD or MFD for the combination has been defined, up to 20 response-evaluable patients with advanced RCC will be enrolled and treated in each of the two study expansion cohorts. The dose expansion phase will confirm the safety and tolerability as well as explore the antitumor activity of axitinib in combination with crizotinib in patients with advanced RCC who are at their first diagnosis (Cohort 1) or have recurred during or after a previous regimen containing a VEGF inhibitor (Cohort 2).

5. ANALYSIS POPULATIONS

5.1. Safety Analysis Set

The safety analysis (SA) set includes all enrolled patients who receive at least one dose of axitinib or crizotinib. This is the primary population for all standard analyses.

5.2. Per Protocol Analysis Set (Evaluable for DLT)

All enrolled patients meeting the inclusion/exclusion criteria, who receive at least one dose of axitinib and crizotinib, and who either experience DLT during the first cycle, or complete the 1st cycle observation period. Patients who are lost to follow-up before receiving at least 75% of the planned first-cycle dose due to reasons other than treatment related adverse events (AEs) are not evaluable for DLT.

5.3. Response Evaluable Analysis Set

All patients who start Cycle 1 (receive at least one dose of axitinib and crizotinib) with an adequate baseline tumor assessment will be considered evaluable for anti-tumor efficacy using standard RECIST 1.1 criteria.

5.4. PK Analysis Set

The PK concentration population is defined as all treated patients who have at least 1 concentration of any of the study drugs.

The PK parameter analysis population is defined as all treated patients who have at least 1 of the PK parameters of interest of any of the study drugs.

The evaluable set for the purpose of PK evaluation will be the treated analysis set of patients who have complete sampling for pharmacokinetic profiles for all drugs.

5.5. Biomarker Analysis Set

The biomarker analysis population is defined separately for blood-based and tumor tissue-based biomarkers. The biomarker analysis population includes all enrolled patients who receive at least one dose of any study drug, and have at least one biomarker parameter from the corresponding assay sample with at least one baseline biomarker measurement. The following are the biomarker analysis sets for the purpose of specific PD biomarker analysis.

Serum Soluble protein biomarker analysis set;
Plasma soluble protein biomarker analysis set;

Tumor miRNA biomarker analysis set;
Tumor IHC analysis set;
Tumor target genes mutation analysis set; and
Tumor gene expression analysis set.

5.6. Treatment Misallocations

Not applicable.

5.7. Protocol Deviations

All deviations will be described when they appear and relate to the statistical analyses or analyses populations.

6. ENDPOINTS AND COVARIATES

6.1. Efficacy Endpoints

Assessment of response will be made using RECIST version 1.1. All response statuses will be derived from lesion measurement data recorded on the electronic case report forms as provided by the investigator.

Tumor assessments will include all known or suspected disease sites. Imaging may include chest, abdomen and pelvis CT or MRI scans; brain CT or MRI scan for patients with known or suspected brain metastases; bone scan for patients with known or suspected bone metastases.

The same imaging technique used to characterize each identified and reported lesion at baseline will be employed in the following tumor assessments.

Patients will be followed and have tumor assessments performed every 8 weeks until disease progression or death, patient refusal, start of another anticancer treatment, or until 18 months from C1D1 of last enrolled patient, whichever occurs first.

- **Objective Response Rate (ORR)** is defined as the percent of patients with confirmed complete response (CR) or confirmed partial response (PR) according to the RECIST 1.1, relative to all response-evaluable patients. Confirmed responses are those that persist on repeat imaging study at least 4 weeks after initial documentation of response. Patients who do not have on-study radiographic tumor re-evaluation or who die, progress, or drop out for any reason prior to reaching a CR or PR will be counted as non-responders in the assessment of ORR. A patient who initially meets the criteria for a PR and then subsequently becomes a confirmed CR will be assigned a best response of CR.

- **Progression Free Survival (PFS)** is defined as the time from the date of first dose of the combination axitinib + crizotinib (C1D1) to the date of the first documentation of objective tumor progression or death on study due to any cause, whichever occurs first. If tumor progression data include more than 1 date, the first date will be used. PFS (in months) will be calculated as (first event date − first dose date +1)/30.44.

  Patients lacking an evaluation of tumor response after the date of the first dose will also have their event time censored on the date of the first dose of the combination (C1D1) unless death occurs prior to 16 weeks (in which case the death is an event).

If patients have at least 1 on-study disease assessment, PFS data will be censored on the date of the last evaluable tumor assessment documenting absence of progressive disease for patients:

- Who are alive and progression free at the time of the analysis (note: summaries will distinguish between subjects still in disease follow-up verses those who have discontinued);
- Who have documentation of disease progression or death after ≥2, consecutive missed tumor assessments (ie, >16 weeks after last tumor assessment);
- Who are given anti-tumor treatment other than the study medication while on study and prior to documented disease progression or death on study.

PFS (months) = [progression/death date – date of first dose +1]/30.44

- **Six-month progression-free survival (6m-PFS)** is summarized as a product limit estimator based on the Kaplan-Meier method to account for censored events.
- **Duration of Response (DR)** is defined as the time from first documentation of objective tumor response (CR or PR) until the first date that recurrent, progressive disease, or death (whichever occurs first) is objectively documented. If tumor progression data include more than 1 date, the first date will be used. Censoring for DR is identical to the censoring rules presented for PFS. DR will only be calculated for the subgroup of patients with a confirmed objective tumor response.

  DR (months) = [progression/death date – first date of OR +1]/30.44

6.2. Safety Endpoints

6.2.1. Dose Limiting Toxicity (DLT) (Primary Endpoint for Dose Finding Cohorts)

Severity of AEs will be graded according to CTCAE v 4.03. AEs meeting one of the definition criteria in protocol Section 3.1.2 occurring in the first cycle of treatment (up to 28 days post C1D1) which are attributable to one or both study drugs will be classified as DLTs.

6.2.2. Adverse Events

AEs will be characterized by type, frequency, severity (as graded by NCI CTCAE version 4.03), timing, seriousness and relationship to study treatment. Adverse events occurring during the 7 day lead in phase for PK will be summarized separately from adverse events reporting after first dose of crizotinib (C1D1).

Baseline signs and symptoms will be recorded at baseline and then reported as adverse events during the trial if they worsen in severity or increase in frequency.

Treatment Emergent Adverse Events

All deaths and serious AEs, regardless of cause, occurring from start of treatment until 28 days after the last dose will be considered treatment-emergent. All other non-serious AEs

occurring after treatment start regardless of cause, will be considered treatment-emergent up until 28 days after the last dose or until start of new anti-cancer treatment, whichever occurs first. Disease progression is not considered a treatment emergent AE unless the patient dies of disease prior to 28 days after the last dose. Events that are continuations of baseline abnormalities are considered treatment emergent AEs only if there is an increase in grade over baseline, or if there is an increase following a decrease during the study.

For subjects with a 7 day lead-in phase for PK, adverse events reported during the lead-in phase will only be considered as treatment emergent for the combination if there is an increase in grade after the first dose of crizotinib (C1D1) or if there is a decrease and then subsequent increase during treatment with the combination.

Treatment Related Adverse Events

Treatment emergent AEs with causality related to treatment, as judged by the investigator, will be considered treatment-related.

6.2.3. Laboratory Abnormalities

Laboratory abnormalities as characterized by type, frequency, severity (as graded by NCI CTCAE v. 4.03) and timing. Abnormalities occurring during the 7 day lead in phase for PK will be summarized separately from abnormalities occurring after first dose of crizotinib (C1D1).

For summaries of the lead-in PK phase and the treatment period starting at C1D1, baseline will be defined as the last evaluation within 28 days prior to the first dose of axitinib and crizotinib, respectively. For laboratory summaries of the entire treatment period (including by cycle), baseline will be defined as the last evaluation within 28 days prior to the first dose of either axitinib or crizotinib, whichever occurs first.

6.2.4. ECG Assessments

The baseline ECG is defined as the most recent ECG(s) prior to (and within the 28 days of) the first dose of either axitinib or crizotinib, whichever occurs first. If the last ECGs prior to first dose of study drug are triplicate ECGs (or only 2), they will be averaged. If the last ECG prior to first dose of study drug is a single measurement, that will be used as baseline.

QT measurements corrected by heart rate (QTc) using Bazett's (QTcB) and Fridericia's (QTcF) methods will be estimated by the Sponsor and used for the data analysis and interpretation.

6.2.5. Opthalmology Examinations

Ocular characteristics, visual acuity and other ophthalmologic endpoints will be collected during fundoscopy and slit lamp examinations.

6.3. PK Endpoints

Pharmacokinetic parameters for crizotinib:

- Cmax, Tmax, AUC0-12, CL/F and V/F as data permit.

Pharmacokinetic parameters for axitinib:

- Cmax, Tmax, AUC0-12, CL/F and V/F as data permit.

6.4. Biomarker Endpoints

Pharmacodynamic biomarkers will be assessed separately for blood and tumor biopsy specimens. These will include measurements of DNA, RNA or protein markers known or suspected to be of relevance to the mechanism of action, the development of resistance, or the identification of those patients who might benefit from treatment with axitinib and crizotinib combination. The assessment will include pre- and post-baseline blood level of plasma and serum biomarkers (e.g. serum soluble protein biomarkers, plasma soluble protein biomarkers), baseline and post-baseline tumor tissue biomarkers (e,g, tumor miRNA, cMET,CD68, CD8, PDL1 tumor IHC).

6.5. Covariates

None.

7. HANDLING OF MISSING VALUES

7.1. Missing Dates

In compliance with Pfizer standards, if the day of the month is missing for any date used in a calculation, the 1st of the month will be used to replace the missing date unless the calculation results in a negative time duration (e.g., date of resolution cannot be prior to date of onset; if replacing resolution date with the 1st of the month results in a negative duration, the resolution date will be set to the onset date.). In this case, the date resulting in 0 time duration will be used. Pfizer standards are also used if both month and day are missing (Jan 1 unless negative time duration). For PFS and DR, if conventions result in a negative duration, durations will be reset to 1 day.

If the start date is missing for an AE, the AE is considered to be treatment emergent unless the collection date is prior to the treatment start date.

7.2. Missing Efficacy Endpoint Values

For all efficacy analyses no values will be imputed for missing data, except as specified in Section 6.1, where for ORR, patients with no post-baseline tumor evaluations will be counted as non-responders.

7.3. ECG

For analyses of ECG results, no values will be imputed for missing data. If one or two of the triplicate measurements for an ECG parameter are missed, the average of the remaining two measurements or the single measurement can be used in the analyses. If all triplicate measurements are missing at a time point for an ECG parameter, no values will be imputed for this time point and no analyses related to this time point will be performed.

7.4. Pharmacokinetics

Concentrations below the limit of quantification:

In all data presentations (except listings), concentrations below the limit of quantification (BLQ) will be set to zero. In listings, BLQ values will be reported as "<LLQ", where LLQ will be replaced with the value for the lower limit of quantification for each drug (axitinib and crizotinib).

Deviations, missing concentrations and anomalous values:

For summary tables and plots of median profiles, appropriate summary statistics will be calculated. Concentrations will be set to missing if one of the following cases is true:

1. A concentration has been reported as ND (i.e., not done) or NS (i.e., no sample),

2. A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the pharmacokineticist.

Note that summary statistics will not be presented at a particular time point if more than 50% of the data are missing. For analysis of PK concentrations, no values will be imputed for missing data.

Pharmacokinetic parameters:

Actual PK sampling times (and where possible the actual dosing information) will be used in the derivation of PK parameters. If a PK parameter cannot be estimated from a subject's concentration data, the parameter will be coded as NC (i.e., not calculated). (Note that NC values will not be generated beyond the day that a subject discontinues).

In summary tables, statistics will be calculated by setting NC values to missing; and statistics will not be presented for a particular treatment if more than 50% of the data are NC. For statistical analyses (i.e., analysis of variance), PK parameters coded as NC will also be set to missing.

If an individual subject has a known biased estimate of a PK parameter (due for example to an unexpected event such as vomiting before all of the drug is absorbed in the body), this will be footnoted in summary tables and will not be included in the calculation of summary statistics or statistical analyses.

For PK analysis, patients will be required to have at least one quantifiable concentration of each drug to be included in the concentration summary. For evaluation of changes in PK of a drug when administered alone vs in combination, only patients with matching pair of PK assessments under both conditions will be included in the PK summary. Patients who have been treated with axitinib and crizotinib for whom drug plasma concentrations (from both PK visits, when administered alone and in combination) are available will be included in average concentration summary.

7.5. Pharmacodynamic Biomarkers

For analysis of pharmacodynamic biomarkers, no values will be imputed for missing data.

Duplicate biomarker (i.e., more than one set of data for a particular visit) is not expected. For continuous data, if duplicate data is received, the results will be averaged and the average value will be used. The average value will be added to the analysis dataset. For non-continuous data, the results will be reviewed by the study team and a representative sample will be selected. The representative sample will be flagged in the analysis dataset.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

8.1.1. Statistical Methods for Dose Escalation: Up-and-Down Matrix Design with the mTPI Method

The mTPI design uses a Bayesian statistics framework and a beta/binomial hierarchical model to compute the posterior probability of three dosing intervals that reflect the relative difference between the toxicity rate of each dose level to the target rate (pT = 0.30). If the toxicity rate of the currently used dose level is far smaller than pT, the mTPI will recommend escalating the dose level; if it is close to pT, the mTPI will recommend continuing at the current dose; if it is far greater than pT, the mTPI will recommend de-escalating the dose level. These rules are conceptually similar to those used by the 3+3 design, except the decisions of an mTPI design are based on posterior probabilities calculated under a coherent probability model.

Being a model-based design, mTPI automatically and appropriately tailors dose-escalation and de-escalation decisions for different trials with different toxicity parameters. More importantly, all the dose-escalation decisions for a given trial can be pre-calculated under the mTPI design and presented in a two-way table (0). Thus, compared to other advanced model-based designs published in the literature, the mTPI design is logistically less complicated and easier to implement. Recently, a Phase 1 trial based on the mTPI design has been published.[4]

Decision rules are based on calculating unit probability mass (UPM) of three dosing intervals corresponding to under, proper, and over dosing in terms of toxicity. Specifically, the underdosing interval is defined as (0; pT-e_1), the over-dosing interval (pT+e_2), and the proper-dosing interval (pT- e_1, pT+ e_2), where e_1 and e_2 are small fractions. Based on the safety profile of axitinib and crizotinib, e_1 is selected as 0.05, and e_2 is selected as 0.00. Therefore, the target dosing interval for the DLT rate is (0.25, 0.30).

The three dosing intervals are associated with three different dose-escalation decisions. The under-dosing interval corresponds to a dose escalation (E), over-dosing corresponds to a dose de-escalation (D), and proper-dosing corresponds to remaining at the current dose (R). Given a dosing interval and a probability distribution, the unit probability mass (UPM) of that dosing interval is defined as the probability of a subject belonging to that dosing interval divided by the length of the dosing interval. The mTPI design calculates the UPMs for the three dosing intervals, and the one with the largest UPM informs the corresponding dose-finding decision, which is the dose level to be used for future patients. For example, if the under-dosing interval has the largest UPM, the decision will be to escalate, and the next cohort of patients

will be treated at the next higher dose level. Ji and collaborators have demonstrated that the decision based on UPM is optimal in that it minimizes a posterior expected loss (ie, minimizes the chance of making a wrong dosing decision)[5].

The dose-finding component of the trial is completed when at least 10 evaluable patients have been treated at the highest dose with DLT rate < 0.33. It is estimated that approximately 25 DLT evaluable patients have been enrolled to reach10-DLT evaluable patients at the estimated MTD.

8.1.2. Methods for Estimating the MTD

The estimated MTD will be the highest tested dose level with a DLT rate < 0.33 in at least 10 DLT evaluable patients at that dose level. We assume that higher doses of either axitinib or crizotinib result in higher toxicity rates. But, due to the relatively low number of patients that may be potentially allocated to any dose combination, this assumption may be violated.

For example, at the end of the escalation phase, the dose combination (crizotinib 250 mg BID, axitinib 3 mg BID) may have a higher proportion of observed toxicities than, say, (crizotinib 250 mg BID, axitinib 5 mg BID), and this variability may be simply related to small cohort size alone. To overcome this potential problem, we use a bivariate isotonic regression to smooth the resulting toxicity surface to a monotonically increasing one. The determination of the MTD contour is accomplished using the Dykstra-Roberston algorithm.[6] Once a monotonically increasing toxicity surface is obtained (either observed or smoothed according to the bivariate isotonic regression algorithm), the MTD combinations closest to the targeted DLT rate of 0.3 but still <0.33 are calculated. Clinical judgment will be exercised in taking forward a particular axitinib + crizotinib regimen to the Expansion Phase cohort(s), in case no clear choice exists between more than 1 competing MTDs. While the limited sample size may result in up to 2 dose combinations of equal potential antitumor activity, under the circumstances of this trial, likely only one will be chosen for the expansion cohort. This decision will be based upon the combination of data related to safety, antitumor activity, and clinical judgment of the Investigators and the Sponsor.

8.1.3. Sample Size Determination

The sample sizes planned for the study arise from logistic feasibility and past experience with Phase 1b studies in oncology and are not entirely driven by statistical considerations. It is expected that approximately 65 patients will be required to achieve all study objectives.

Due to the dynamic nature of the Bayesian allocation procedure, the sample size of the up-and-down matrix design using the mTPI approach cannot be determined in advance. It is estimated that 25 DLT evaluable patients will be enrolled in the dose escalation stage in order to have a reliable and accurate estimate of the MTD. In addition, there will be Dose Expansion Phase cohorts to characterize safety, biomarkers, and efficacy in terms of probability (p) of achieving an event of interest including, but not limited to, objective response (OR). The goal will be to estimate proportions of such patients with the standard error (SE) of not greater than 0.12, i.e., by definition,

$$SE = \sqrt{\frac{p(1-p)}{n}} \leq \frac{1}{2\sqrt{n}}$$

Therefore, a sample of twenty patients (*n=20*) per Dose Expansion Phase cohort will allow estimation of the probability of achieving an event of interest with the standard error <0.12.

8.2. Statistical Methods for Different Types of Endpoints

Listings and standard summary statistics will be used to analyze the study.

Analysis of Time-to-Event Endpoints

Time-to-event endpoints will be summarized using the Kaplan-Meier method and displayed graphically when appropriate. Medians and two-sided 95% confidence intervals will be provided.

Analysis of Binary Endpoints

Binary endpoints will be summarized by percentage rates along with the 95% confidence intervals using an exact method.

Analysis of Continuous Endpoints

Continuous endpoints will be summarized by descriptive statistics, including the mean, standard deviation, median, minimum, and maximum values. Biomarker analytes will also include %CV (as appropriate).

Analysis of Categorical Endpoints

The number and percentage of patients in each category will be provided for categorical variable.

8.3. Statistical Analyses

8.3.1. Standard Analyses

- **Study Conduct and Patient Disposition** - an accounting of the study patients in the SA population will be provided. Enrolled patients not meeting the eligibility criteria will be listed. Patients not completing the study will be listed along with the reason for their discontinuation. Reasons for discontinuation will be summarized.

- **Baseline Characteristics** – for patients in the SA population, characteristics such as age, height, weight, race, ethnicity, diagnosis, performance status, and medical history at study entry will be summarized in frequency tables, and descriptive statistics will be provided for quantitative variables.

- **Treatment Administration/Compliance** – Administration of study medication will be presented for the SA population, by medication administered and will be described in terms of the duration of treatment in days (mean, median and range), number of cycles administered (mean, median and range), dose intensity, dose modifications, dose interruptions, dose delays.

8.3.2. Analysis of Efficacy Endpoints

PFS and DR will be summarized using the Kaplan-Meier method and displayed graphically. The median event time and 2-sided 95% confidence interval (CI), calculated using the Brookmeyer-Crowley method, for the median will be provided for each endpoint.

Six-month progression-free survival (6m-PFS) will be summarized as a product limit estimator based on the Kaplan-Meier method to account for censored events, together with the corresponding 2-sided 95% CI. The 2-sided 95% CI for the log[-log(6-month PFS probability)] will be calculated using normal approximation and then back transformed to give the CI for the 6-month PFS rate itself. Analysis result will be included in the table for the PFS analysis.

The ORR will be summarized along with the corresponding exact 2-sided 95% CI calculated using a method based on the F distribution. If a patient has not achieved an objective response, but remain stable for at least 8 weeks, then the best overall response for such a patient will be SD. Patients will be followed and have tumor assessments performed every 8 weeks until disease progression or death, patient refusal, start of another anticancer treatment, or until 18 months from C1D1 of last enrolled patient, whichever occurs first.

Tumor Response will be presented in the form of patient data listings that include, but are not limited to, lesion type (target/non-target), received (maximum) dose, overall tumor response at each visit, and best overall response. In addition, progression date, death date, date of first response and last tumor assessment date, dates of first dose and last dose will be listed, together with DR and PFS.

Listings or/and tables (when applicable) will be sorted by treatment cohort.

Analysis Sets

- Response-evaluable population for Tumor Response (dose escalation and dose expansion cohorts) and PFS (dose expansion cohorts).

- Patients with an overall objective response of CR or PR in the response-evaluable population for Duration of Tumor Response (dose expansion cohorts only).

8.3.3. Analysis of Safety Endpoints

Analysis Set

- Summaries and analyses of the primary safety endpoint will be based on the per protocol analysis set. Summaries and analyses of all other safety parameters will include all patients in the Safety Analysis Set.

- Adverse events and laboratory tests will be summarized for the following study periods: entire study period and by cycle (Cycle 1 and Cycles beyond 1) in the Safety Analysis Set; the PK lead-in period for those patients receiving at least one dose of axitinib during the 7 day PK lead-in period; and the treatment period starting at C1D1 for patients receiving the combination treatment (Dose escalation cohorts and Dose expansion cohort 1 only).

8.3.3.1. Analysis of Primary Endpoint

Dose Limiting Toxicity (DLT) is the primary endpoint of the the study. The occurrence of DLTs observed in the dosing cohorts is used to estimate the MTD as described in the protocol. AEs constituting DLTs will be listed per dose level.

DLT-related listings will be produced by dose

- Patient ID.
- Dose, Cycle and Date at which DLT occurred.
- Time from treatment start to onset of DLT.
- Time to resolution of DLT to Grade 1 or baseline.
- Dose interruption (yes, no).
- Time to resumption of treatment.
- DLT term.
- Action(s) taken due to DLT (stopped temporarily, permanently discontinued, no action taken, etc).

8.3.3.2. Analysis of Secondary Safety Endpoint

Adverse Events

AEs will be graded by the Investigator according to the Common Terminology Criteria for Adverse Events (NCI CTCAE) version 4.03 and coded using the Medical Dictionary for Regulatory Activities (MedDRA). The focus of AE summaries will be on treatment-emergent AEs The number and percentage of patients who experienced any AE, serious AE (SAE), treatment-related AE, and treatment-related SAE will be summarized according to worst toxicity grades. The summaries will present AEs on the entire study period, separately for the lead-in PK period and by cycle (Cycle 1 and Cycles beyond 1), as well as for the entire treatment period starting at C1D1 (first dose of crizotinib) for patients receiving the combination treatment in the Dose escalation cohorts and Dose expansion cohort 1.

For AE summaries of the treatment period starting at C1D1, treatment-emergent will be defined as an adverse event with an initial onset or increasing severity after the first dose of crizotinib. For all other AE summaries, treatment-emergent will be defined as an adverse event with an initial onset or increasing severity after the first dose of any study medication (axitinib, in the case of the lead-in PK period).

- An overall summary of AEs will be provided. The number and percentage of patients who experienced any AE, who experienced any SAE, who experienced any treatment-related AE, who experienced any treatment-related SAE, and who

discontinued because of an AE will be presented. Treatment-related AEs are those judged by the investigator to be at least possibly related to the study drug (with a cause related to the study drug indicated on the CRF).

- All treatment-emergent AEs will be summarized by MedDRA SOC and preferred term. A summary of AEs by preferred term and maximum CTCAE grade will be presented. A summary of AEs by preferred term and maximum CTCAE grade group (Grade 1-2, Grade 3-4, and Grade 5) will also be presented. Treatment-emergent AEs will also be summarized for each group and pooled across the groups.

- Treatment-related AEs will be summarized by MedDRA SOC and preferred term separately for axitinib, crizotinib and both. A summary of treatment-related AEs by preferred term and maximum CTCAE grade will be presented. A summary of treatment related AEs by preferred term and maximum CTCAE grade group (Grade 1-2 vs. Grade 3-5) will also be presented. Treatment-related AEs will also be summarized for each group and pooled across the groups.

- On-treatment deaths will be summarized by cause of death. Deaths that occurred within 28 days after the last dose of study treatment are defined as on treatment deaths. Death data will also be listed.

- Patients who withdraw from study treatment because of an AE will be listed separately for axitinib and crizotinib. Patient discontinuation will be determined from the end of treatment (EOT) evaluation (where reason for termination is "Adverse Event") and the specific AE(s) will be determined from the AE CRF page (where action taken is "Withdrawn from Treatment").

- SAEs and treatment-related SAEs will be summarized by MedDRA SOC and preferred term. Patients who experienced a SAE will be listed.

For all safety analyses, listings or/and tables (when applicable) will be sorted by treatment cohort. Only descriptive methods will be used without any formal statistical test.

Laboratory Tests

The number and percentage of patients who experienced laboratory test abnormalities will be summarized according to worst toxicity grade observed for each lab assay. The analyses will summarize laboratory tests on the entire treatment period, separately for the lead-in PK period and by cycle (Cycle 1, and Cycles beyond 1), as well as for the entire treatment period starting at C1D1 (first dose of crizotinib) for patients receiving combination therapy treatment (Dose escalation cohorts and Dose expansion cohort 1 only).

For laboratory tests without CTC grade definitions, results will be categorized as normal, abnormal or not done.

- **Hematology –** Descriptive statistics will be provided for each test result and for change from baseline by visit. Hematology results will be graded according to the NCI CTCAE Version 4.03. A summary of maximum CTCAE grade as well as shift

summary of baseline grade by maximum CTCAE grade, cycle, and dose will be presented. Patients who developed a grade 3 or greater toxicity will also be listed.

- **Biochemistry -** Descriptive statistics will be provided for each parameter result and for change from baseline by visit. Biochemistry results will be graded according to the NCI CTCAE version 4.03. A summary of maximum CTCAE grade as well as shift summary of baseline grade by maximum CTCAE grade, cycle, and dose will be presented. Patients who developed a grade 3 or greater toxicity will also be listed.

- **Urine** - Descriptive statistics will be provided for urine protein and blood results and for change from baseline by visit. Urine protein and blood data will also be listed and patients with urine protein ≥2+ by semiquantitative method (e.g., dipstick) will have their 24 hour urine collection protein levels listed as well. A shift summary of baseline value by maximum value, cycle and dose will be presented for urine protein.

- **Other Laboratory Tests** – Individual patient test results will be listed.

ECOG Performance Status

ECOG performance status data will be summarized with a shift table of the baseline and the worst on study status. Baseline will be defined as the last evaluation within 28 days prior to the first dose of either axitinib or crizotinib, whichever occurs first.

Vital Signs

Summaries and listings will be presented for blood pressure, body weight and pulse rate.

In addition, the baseline and the change from baseline in blood pressure and pulse rate will be summarized using descriptive statistics by visit. Baseline will be defined as the last evaluation within 28 days prior to the first dose of either axitinib or crizotinib, whichever occurs first.

ECG

The analysis of ECG results will be based on the SA set for patients having baseline and on-treatment ECG data. The triplicate ECG data will be averaged and all summary statistics and data presentations will use the averaged data. Any data obtained from ECGs repeated for safety reasons after the nominal time-points will not be averaged along with the preceding triplicates. Any data obtained from hand-reading of abnormal ECGs by a cardiologist (reported as unplanned ECGs in eCRF) will not be included in the summary tables but will be averaged separately and provided in the data listings along with the machine-calculated values.

QT measurements corrected by heart rate will be estimated by the Sponsor and used for the data analysis and interpretation. The commonly used Bazett's correction (QTcB) and Fridericia's correction (QTcF), and study specific correction, if necessary, will be applied.

Individual absolute values and changes from baseline in QT, heart rate, QTc (including but not limited to QTcB and QTcF), PR and QRS will be summarized by cohort and nominal post-dose time points using descriptive statistics according to Sponsor reporting standards. If data permit, mean changes from baseline in heart rate and QTc (all evaluated corrections) will be plotted against time post-dose by cohort.

Maximum increases from baseline for QT, Heart Rate, QTc, PR and QRS for individual will be summarized by treatment according to Sponsor reporting standards.

Categorical analysis of ECG data will be conducted and summarized as follows. All planned and unplanned post-dose time points will be counted in these categorical summaries. All values meeting the criteria of potential clinical concern will be listed.

Categorical analyses of the QTcF/QTcB data will be conducted and summarized as follows:

1. The number and percentage of patients with maximum increase from baseline in QTcF/QTcB (<30, 30- 60, and ≥60 ms);
2. The number of and percentage patients with maximum post-dose QTcF/QTcB (<450, 450-480, 481-500, and ≥501 ms);
3. PR changes from baseline ≥50% if absolute baseline value was <200 ms, and ≥25% if absolute baseline value was ≥200 ms;
4. QRS changes from baseline ≥50% if absolute baseline value was <100 ms, and ≥25% if absolute baseline value was ≥100 ms.

Shift tables will be provided for baseline vs. worst on-treatment QTc using maximum CTCAE version 4.03. Patients with QTc values >500 ms or maximum increase from baseline ≥60 ms will be listed. In addition, the baseline and the change from baseline of ECG interval and heart rate will be summarized using descriptive statistics cohort and by visit.

Concomitant Medications and Non-drug Procedure/Treatments

All drug medications will be coded by the World Health Organization (WHO) medical dictionary. Non-drug procedure/treatments will be coded by the MedDRA dictionary. All medications with a start date prior to lead-in D1 (for patients with a lead-in period) or prior to C1D1 (for patients with no lead-in period) are considered previous medication. All ongoing medications at the end of study or medications with a stop date on or after lead-in D1 (for patients with a lead-in period) or on or after C1D1 (for patients with no lead-in period) are considered concomitant medication. If a medication satisfies both the definition of previous and concomitant medication, it will be considered both previous and concomitant medication.

The number of subjects with any concomitant drug/non-drug treatment will be summarized. Listings of prior and concomitant drug/non-drug treatment will be provided separately. If any prior or concurrent surgery or radiation therapy is given, these data will be listed for each patient. Furthermore, listings of previous systemic therapy for primary diagnosis will be provided.

Ophthalmologic Data

Best corrected visual acuity (Snellen fraction), biomicroscopic, and ophthalmoscopic findings will be recorded at screening and any change in baseline from screening will be summarized/listed for all patients who receive a dose of either axitinib or crizotinib, with a baseline assessment and at least one post-baseline assessment.

For those patients selected for additional testing, the refractive error (sphere, cylinder, axis), pupillary size (millimeters), optical coherence tomographic central retinal subfield thickness (microns), and intraocular pressure (mm Hg) at screening and their changes from screening will be quantitatively summarized/listed. The qualitative findings of fundus photography and optical coherence tomography at screening and their changes from screening will also be summarized/listed.

For the baseline screening results, percentages of patients falling into each category of the examination status (normal, mild, moderate, or severe) will be summarized/listed for each structure by eye. For post-baseline results, percentages of patients falling into each category of the examination status (new findings/worsening of finding, no change, improvement of finding, etc.) will be summarized/listed for each eye structure**.**

Data Safety Monitoring Committee

A Data Safety Monitoring Committee will not be established for the study.

8.3.4. Analysis of Pharmacokinetics

Analysis Set: PK

8.3.4.1. Pharmacokinetic Analysis of Crizotinib and Axitinib

Standard PK parameters for axitinib (Cmax, Tmax, AUC0-last, AUC0-12, AUC_{0-24}, CL/F, Vz/F) and plasma elimination half life ($t_{1/2}$). and crizotinib (Cmax, C_{min},Tmax, AUC0-last, AUC0-12, CL/F) will be estimated using non-compartmental methods.

Presentation of pharmacokinetic data will include:

- Descriptive statistics (n, mean, SD, %CV, median, minimum, maximum) of plasma concentrations for axitinib and crizotinib will be presented in tabular form by treatment cohort, dose level, cycle, day and nominal time.

- Linear-linear and log-linear plots of mean and median plasma concentrations by nominal time for axitinib and crizotinib will be presented for PK sampling days by treatment cohort, cycle, and study day. Similar plots will be presented for axitinib and crizotinib for each individual patient concentrations. Only patients who have matched pairs of PK collections available on both planned treatments (when administered alone and in combination) will be included in the axitinib plasma concentration descriptive summary and median concentration profiles. Patients who have undergone intra-patient dose reduction or escalation will be excluded from the median plasma concentration-time plots.

- Pharmacokinetic parameters for axitinib and crizotinib will be listed and summarized by treatment cohort/dose level, cycle and study day using descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% confidence interval). Only patients who have a matched pair of estimated PK parameters available on both planned treatments (when administered alone and in combination) will be included in the axitinib PK parameter summary tables. PK parameters with zero values will be excluded from the calculation of geometric means and its associated %CV. If an intra patient dose escalation or reduction occurs, dose-dependent PK parameters (AUC and C_{max}) for that patient may be dose-normalized when it is known that the drug exhibits linear PK within the dose range and other PK parameters will be reported as estimated; or may only be included in descriptive statistics and summary plots up to the time of the dose change.

- Box plots for AUC and C_{max} for each drug (during treatments when given alone and in combination) will be generated. Individual data points, the geometric mean and the median of the parameter in each treatment will be overlaid on the box plots. If a cohort has limited evaluable PK data (n<4), matchstick plots showing changes in AUC and C_{max} for each drug (during treatments when given alone and in combination) in individual patients will then be generated. The geometric mean of the parameter in each treatment will be overlaid in the plots.

- Trough concentrations for crizotinib will be plotted for each treatment cohort using a box-whisker plot by cycle and day within cycle in order to assess the attainment of steady-state.

8.3.4.2. Effect of Crizotinib on Axitinib Pharmacokinetics

The effect of repeated crizotinib dosing on axitinib PK will be evaluated using AUC_{0-12} of axitinib on Lead-in Day 7 and Cycle 1 Day 15, respectively, as the primary PK parameter. Ninety-percent confidence interval for the ratio of geometric means of AUC_{0-12} and Cmax (axitinib in presence of crizotinib/axitinib alone) will be computed to assess the magnitude of the effect.

8.3.4.3. Population Pharmacokinetic Analysis or PK/PD Modeling

Mechanism-based or semi-mechanistic sequential pharmacokinetic-pharmacodynamic models may be developed using NONMEM® to explore any relationships between plasma drug concentrations and selected safety, biomarker, and efficacy endpoints.

The results of these analyses, if performed, will be reported separately.

8.3.5. Analysis of Biomarker Endpoints

Pharmacodynamic biomarkers will be summarized by cohort, Naïve (1L) and Later-Line (2L), using the appropriate biomarker analysis set. Summaries of baseline and ratio to baseline values at post-treatment visits will be provided using N, mean, standard deviation, median, % CV (as appropriate) and minimum/maximum for soluble protein biomarkers (including sMET from plasma and Ang-2, VEGFR2, IL6, etc. from serum), and tumor RNA measurements. Baseline tumor IHC scoring (e.g., H-score, % positive cells; number of positive cell per

surface area; immunoscore, where applicable) for cMET, CD68, CD8, PDL1 expression will be summarized for expansion cohort 1 and 2. Tumor target gene (e.g., might include PBRM1, BAP1, SETD2, etc) mutations will be summarized with number and percentage at baseline and end of treatment for dose escalation and expansion cohort 1 and 2. Tumor gene expression profiling will be summarized at baseline and end of treatment, likely for dose escalation and expansion cohort 1 and 2. Comparisons between the two cohorts will be analyzed using the Wilcoxon Rank Sum test for biomarker levels of Baseline tumor IHC scorings, Baseline and End of Treatment tumor target genes, and gene expression profiling. P-values will be provided if N≥5 in either of the cohort.

Correlations of biomarker results with measures of anti-tumor efficacy will be examined. Summary of level of biomarkers at baseline and/or ratio of values to baseline versus ORR category will be made by timepoint and cohort. Summary of PFS and DR will be provided after stratification by a preset cutpoint of biomarker values at baseline and/or ratio of values to baseline by timepoint and treatment cohort. The median PFS and DR will be estimated using Kaplan-Meier method together with the corresponding 2-sided 95% confidence interval; p-values will not be displayed when N<10 in either group. The estimated hazard ratio and its 2-sided 95% confidence intervals (CI) will be reported for comparison between groups. If there are too few events to meaningfully interpret the Kaplan-Meier analysis, neither p-value nor hazard ratio will be reported.

Graphical display will be provided for level of biomarkers such as serum and plasma soluble protein levels at baseline and/or ratio of values to baseline at each timepoint and by cohort. Box and Whisker plots will be produced for level of biomarkers at baseline versus ORR category by cohort. Kaplan-Meier plots will be produced for PFS and DR after stratification by < or >= median levels of biomarkers at baseline by cohort.

All biomarker data will be listed.

8.4. Summary of Secondary Efficacy Analyses

Endpoint	Analysis Population	Statistical Method	Missing Data	Analysis Type/Timing
ORR	Response Evaluable	Exact method based on F-distribution (95% CI) (See 8.3.2)	See Section 6.1	Secondary analysis
DR	Subgroup of pts with OR from the Response Evaluable	K-M method (median and 95% CI) (See 8.3.2)	See Section 6.1	Secondary analysis
PFS	Response Evaluable	K-M method (median, 6m-PFS and 95% CIs) (See 8.3.2)	Censor patients on the day of the last evaluable tumor assessment documenting absence of disease progression for … (See Section 6.1)	Secondary analysis

CI: Confidence intervals; DR: Duration of Response; K-M: Kaplan-Meier; OR: Objective Response; ORR: Objective Response Rate; PFS: Progression-free Survival; SA: Safety Analysis.

9. REFERENCES

1. Brookmeyer R, Crowley JJ. A confidence interval for the median survival time. Biometrics. 38: 29-41, 1982.

2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 53: 457-81, 1958.

3. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan;45(2):228-47.

4. Fanale M et al. Phase I study of bortezomib plus ICE (BICE) for the treatment of relapsed/refractory Hodgkin lymphoma. British Journal of Haematology, 154:284 286, 2011.

5. Ji Y et al. A modified toxicity probability interval method for dose-finding trials. Clinical Trials 2010; 7:653 663.

6. Dykstra R, Robertson T. An algorithm for isotonic regression for two or more independent variables. *Ann Stat.* 1982;10:708-716.

10. APPENDICES

Appendix 1. Programming Specifications for AE Analyses

a. Time to AE onset

1. Definition

Time to AE onset (days) will be calculated as first *AE start date – first dose date* +1. The definition and calculations are similar for time to Grade 3/4 AE onset.

AE start date

The Date of Onset for the first occurrence of the AE based on the Log AE CRF page.

First dose date

For AE summaries of the lead-in PK period: The date of the first dose taken from the Axitinib Dosing CRF page.

For AE summaries of the treatment period starting at C1D1 (Dose escalation cohorts and Dose expansion cohort 1): The date of the first dose taken from the Crizotinib Dosing CRF page.

For AE summaries of the entire treatment period (including by cycle): The date of the first dose taken from either the Axitinib or Crizotinib Dosing CRF page, whichever occurs first.

b. Duration of AE

1. Definition

Duration of AE (days) is defined as the cumulative duration across episodes of the specific AE (by preferred terms) where duration for each episode is calculated as *AE end date – AE start date* +1 excluding any overlap. Duration of AE is defined for subjects with the AE.

AE start date

The Date of Onset based on the Log AE CRF page.

AE end date

The Date Resolved based on the Log AE CRF page.

2. Censoring

AE resolution is considered an event (censoring variable=1). If a subject has an AE that was ongoing (does not have to be the last AE) at the time of analysis, the time is

censored (censoring variable=0) at the last available on treatment visit date. Subjects who die prior to resolution of the AE will be censored at the *date of death*. If the date of death is the same as the date of the resolution of the AE, the subject will be censored at that date (i.e. resolution will not be considered an event) and only if the AE is the AE that resulted in death will it be counted as an event.

Date of death

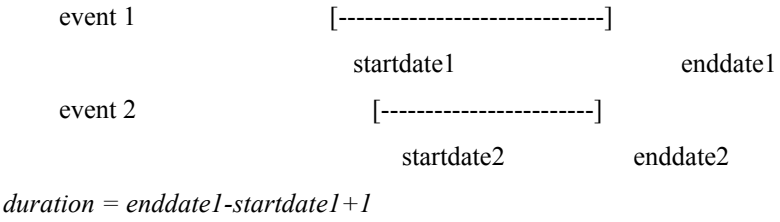
Death date is based on the Notice of Death CRF page.

3. Clustered Events

For clustered events, a patient could have multiple events in the cluster which may overlap. In this case, AE duration will be summed across all events in the cluster accounting for the overlap (i.e. overlapping periods between events in the same cluster are not double-counted). Lags between events in the same cluster are not included in the duration.

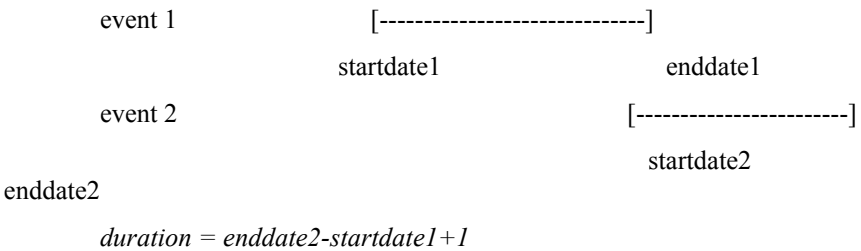
The following scenarios provide examples of the calculation for 2 events in the same cluster. The extension to 3 or more events of the same cluster is similar.

- **TWO EVENTS OF THE SAME CLUSTER WHERE ONE EVENT COMPLETELY CONTAINS THE OTHER EVENT**



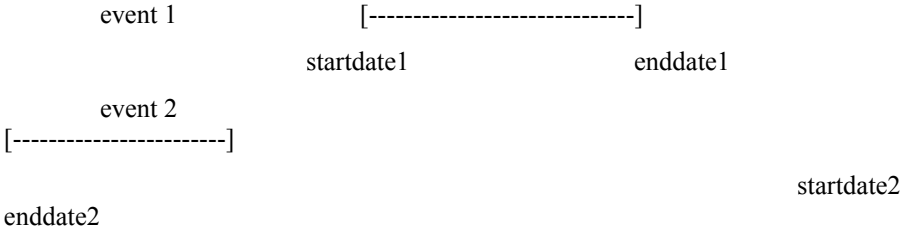
duration = enddate1-startdate1+1

- **CERTAIN PORTIONS OF TWO EVENTS IN THE SAME CLUSTER OVERLAP**



duration = enddate2-startdate1+1

- **TWO EVENTS OF THE SAME CLUSTER ARE CONTIGUOUS TO EACH OTHER**



duration = enddate2-startdate1+1

- **TWO EVENTS OF THE SAME CLUSTER ARE NON-OVERLAPPING**

event 1 [----------]

startdate1 enddate1

event 2
[----------]

startdate2 enddate2

duration = (enddate1-startdate1+1) + (enddate2-startdate2+1)

c. AE Prevalence

1. Definition

AE prevalence is defined as the number of patients with an AE in a particular time period (including both new cases with an onset date during the specified time period AND cases with an AE continued from a previous time period) divided by the number of patients at risk during the specified time period. The number of patients at risk includes all subjects except those who either have discontinued or died prior to the specified time period.

AE prevalence will be presented for the entire study period, separately for the lead-in PK period and by cycle (Cycle 1 and Cycles beyond 1), as well as for the entire treatment period starting at C1D1 (Dose escalation cohorts and Dose expansion cohort 1).

2. Assumptions

For AE summaries of the study period starting at C1D1 the following apply:

- A patient is counted in the numerator if the patient has an onset date of the AE during Cycle 1 OR for patients with a 7 day lead-in PK period there is an increase in grade of the AE after the first dose of crizotinib (C1D1) or if there is a decrease and then subsequent increase during treatment with the combination.

- The denominator will include subjects who have received at least one dose of crizotinib.

For AE summaries of the entire study period including by cycle (Lead-in PK period, Cycle 1 and Cycles beyond 1) the following apply:

- Patients are counted for an AE in each cycle up until the cycle where the AE resolved. Thus, the calculation conservatively assumes that if the AE resolved in a cycle, it resolved at the end of the cycle.

- For Lead-in PK period, a patient is counted in the numerator if the patient has an onset date of the AE during the Lead-in period.

- For Cycle 1, a patient is counted in the numerator if the patient has an onset date of the AE during Cycle 1 OR an onset date during the Lead-in PK phase that is still ongoing in Cycle 1 (did not have a resolution date in the Lead-in phase).

- For Cycles >1, a patient is counted in the numerator if the patient has an onset date in Cycles >1 OR an onset date in a previous cycle that is still ongoing (did not have a resolution date in Cycle 1 or earlier).

- The denominator for a particular time period will include subjects who are at risk prior to the time period. The number at risk includes all subjects except those who

either have discontinued (based on the 'Subject Summary' CRF) or died prior to the specified time period (i.e. Death Date based on 'Notice of Death' CRF is prior to start of time period).

Appendix 2. Study Specific Information for Efficacy

- **Baseline**: is defined as the last observation within 28 days prior to the first dose of any study treatment (first dose of axitinib for patients receiving the lead-in PK dose or first dose of either drug for patients not receiving the lead-in PK dose).

- **Adequate Baseline:**

    - Baseline tumor evaluations must be performed within 4 weeks (28 days) prior to the first dose of any study treatment (first dose of axitinib for patients receiving the lead-in PK dose or first dose of either drug for patients not receiving the lead-in PK dose);

    - Presence of at least one measurable lesion per RECIST version 1.1 for patients in the dose expansion cohorts;

    - All lesions recorded at baseline must have an associated status recorded on the CRF;

    - Baseline lesions must be assessed with an acceptable method that includes: Conventional CT Scan, Spiral CT Scan, X-ray, MRI, Physical Exam, Bone Scan and Other. Note: If based on data review "unacceptable" methods (e.g. ultrasound, etc) are noted under "Other", then this category will not be considered acceptable (on a case by case basis).

- **"On-study" period for efficacy:** is defined as the period from the date of the first dose of crizotinib until subject death, progression of disease, subject no longer willing to participate, start of other anti-cancer treatment, whichever is earlier.

- **Subsequent anti-tumor treatment:** includes any systemic anticancer therapy (other than study medication), radiation to target lesions, and surgery for removal (resected or partially resected) of target lesions.

Appendix 3. RECIST 1.1

The determination of antitumor efficacy during this study will be based on objective tumor assessments made according to the RECIST system of unidimensional evaluation.

Measurability of Tumor Lesions

At baseline, individual tumor lesions will be categorized by the Investigator as either measurable or non-measurable by the RECIST criteria as described below.

Measurable:

Tumor lesion: Must be accurately measured in at least one dimension (longest diameter in the plane of measurement is to be recorded) with a minimum size of:

- 10 mm by CT scan (CT scan slice thickness no greater than 5 mm);
- 10 mm caliper measurement by clinical exam (lesions which cannot be accurately measured with calipers should be recorded as non-measurable).

Malignant lymph nodes: To be considered pathologically enlarged and measurable, a lymph node must be ≥15mm in short axis when assessed by CT scan (CT scan slice thickness recommended to be no greater than 5 mm). At baseline and in follow-up, only the short axis will be measured and followed.

Non-Measurable: All other lesions, including small lesions (longest diameter <10mm or pathological lymph nodes with ≥10 mm to <15 mm short axis) as well as truly non-measurable lesions. Lesions considered truly non-measurable include: leptomeningeal disease, ascites, pleural or pericardial effusion, inflammatory breast disease, lymphangitic involvement of skin or lung, abdominal masses/abdominal organomegaly identified by physical exam that is not measurable by reproducible imaging techniques.

NOTE: If measurable disease is restricted to a solitary lesion, its neoplastic nature should be confirmed by cytology/histology.

Recording Tumor Measurements

All measurable lesions up to a maximum of 2 lesions per organ and 5 lesions in total representative of all involved organs should be identified as **target lesions** and measured and recorded at baseline and at the stipulated intervals during treatment. Target lesions should be selected on the basis of their size (lesion with the longest diameters) and their suitability for accurate repetitive measurements (either by imaging techniques or clinically).

The longest diameter will be recorded for each target lesion. The sum of the longest diameter for all target lesions will be calculated and recorded as the baseline sum longest diameter to be used as reference to further characterize the objective tumor response of the measurable dimension of the disease during treatment. All measurements should be performed using a caliper or ruler and should be recorded in metric notation in centimeters.

All other lesions (or sites of disease) should be identified as **non-target lesions** and should also be recorded at baseline. Measurements are not required and these lesions should be followed as "present" or "absent."

Techniques for Assessing Measurable Disease

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at screening and during follow-up. Imaging-based evaluation is preferred to evaluation by clinical (physical) examination when both methods have been used to assess the antitumor effect of a treatment.

Definitions of Tumor Response

Target Lesions

Complete response (CR) is defined as the 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) is defined as a ≥30% decrease in the sum of the longest dimensions of the target lesions taking as a reference the baseline sum longest dimensions.

Progressive disease (PD) is defined as a ≥20% increase in the sum of the longest dimensions of the target lesions taking as a reference the smallest sum of the longest dimensions recorded since the treatment started, or the appearance of one or more new lesions. In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm.

Stable disease (SD) is defined as neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD taking as a reference the smallest sum of the longest dimensions since the treatment started.

Non-Target Lesions

Complete response (CR) is defined as the disappearance of all non-target lesions. All lymph nodes must be non-pathological in size (<10mm short axis).

Non-CR/Non-PD is defined as a persistence of ≥1 non-target lesions.

Progressive disease (PD) is defined as unequivocal progression of existing non-target lesions, or the appearance of ≥1 new lesion.

The cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment when the measurable tumor has met criteria for response or stable disease is mandatory to differentiate between response or stable disease and progressive disease.

Confirmation of Tumor Response

To be assigned a status of PR or CR, changes in tumor measurements in patients with responding tumors must be confirmed by repeat studies that should be performed ≥4 weeks after the criteria for response are first met. In the case of SD, follow-up measurements must have met the SD criteria at least once after study entry at a minimum interval of 8 weeks.

Determination of Tumor Response by the RECIST Criteria

When both target and non-target lesions are present, individual assessments will be recorded separately. Determination of tumor response at each assessment is summarized in the following table.

Response Evaluation Criteria in Solid Tumors

Target Lesions[1]	Non-Target Lesions[2]	New Lesions[3]	Tumor 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
PD	Any response	Yes or No	PD
Any response	PD	Yes or No	PD
Any response	Any response	Yes	PD

1 Measurable lesions only.
[2] May include measurable lesions not followed as target lesions or non-measurable lesions.
[3] Measurable or non-measurable lesions

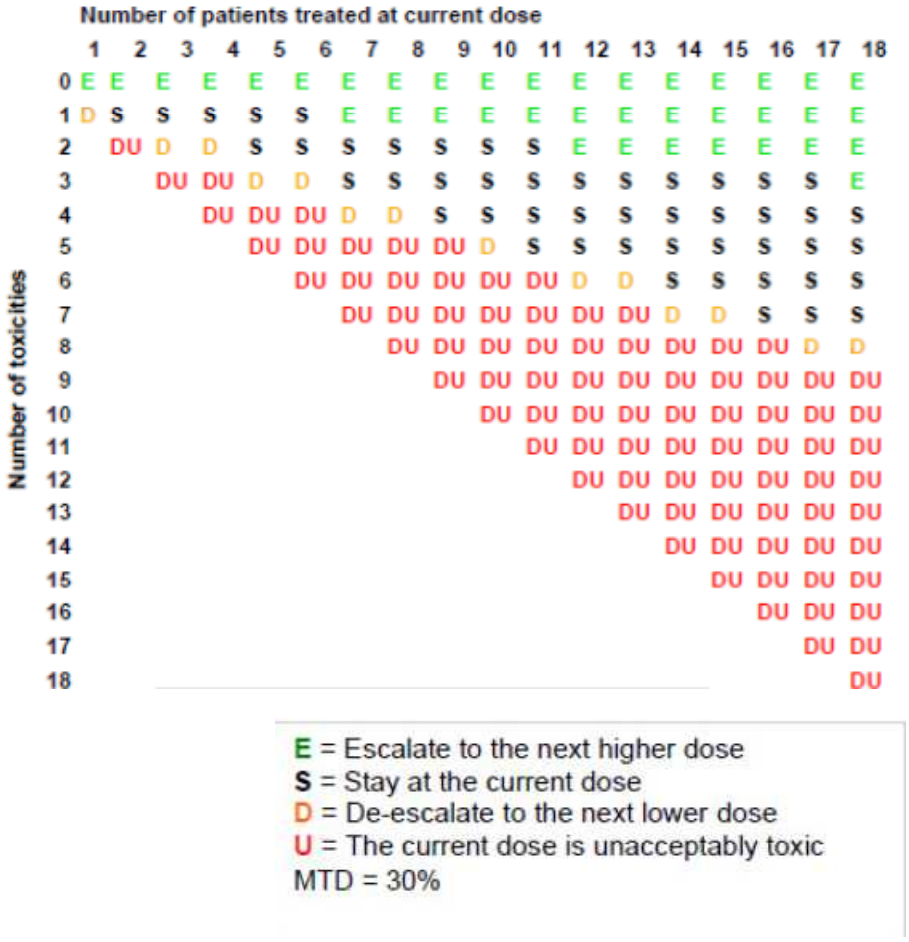
Determination of Best Overall Response

The best overall response is the best response recorded from the start of the treatment until disease progression/recurrence (taking as reference for progressive disease the smallest measurements recorded since the treatment started). For CR and PR, the patient's best response assignment will depend on the achievement of both measurement and confirmation criteria. In the case of SD, follow-up measurements must have met the SD criteria at least once after study entry at a minimum interval of 8 weeks.

NOTE: Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as "symptomatic deterioration." Every effort should be made to document the objective progression even after discontinuation of treatment. It should also be noted that a tumor marker increase does not constitute adequate objective evidence of tumor progression. However, such a tumor marker increase should prompt a repeat radiographic evaluation to document whether or not objective tumor progression has occurred.

In some circumstances, it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends upon this determination, it is recommended that the residual lesion be investigated by fine needle aspirate or biopsy before confirming the complete response status.

Appendix 4. Detailed Dose Escalation/De-Escalation Scheme



Escalation/De-escalation algorithms for total number of patients treated at the current dose level (current and previous cohorts)

- With 3 patients treated at current dose level
  - 0 DLT -> escalate
  - 1 DLT -> remain at the same dose
  - 2 DLTs -> de-escalate
  - 3 DLTs -> de-escalate and consider current dose as intolerable
- With 4 patients treated at current dose level
  - 0 DLT -> escalate
  - 1 DLTs -> remain at the same dose
  - 2 DLTs -> de-escalate

    - 3-4 DLTs -> de-escalate and consider current dose as intolerable
- With 5 patients treated at current dose level
    - 0 DLT -> escalate
    - 1-2 DLTs -> remain at the same dose
    - 3 DLTs -> de-escalate
    - 4-5 DLTs -> de-escalate and consider current dose as intolerable
- With 6 patients treated at current dose level
    - 0 DLT -> escalate
    - 1-2 DLTs -> remain at the same dose
    - 3 DLTs -> de-escalate
    - 4-6 DLTs -> de-escalate and consider current dose as intolerable
- With 7 patients treated at current dose level
    - 0-1 DLT -> escalate
    - 2-3 DLTs -> remain at the same dose
    - 4 DLTs -> de-escalate
    - 5-7 DLTs -> de-escalate and consider current dose as intolerable
- With 8 patients treated at current dose level
    - 0-1 DLT -> escalate
    - 2-3 DLTs -> remain at the same dose
    - 4 DLTs -> de-escalate
    - 5-8 DLTs -> de-escalate and consider current dose as intolerable
- With 9 patients treated at current dose level
    - 0-1 DLT -> escalate
    - 2-4 DLTs -> remain at the same dose
    - 5-9 DLTs -> de-escalate and consider current dose as intolerable
- With 10 patients treated at current dose level
    - 0-1 DLT -> escalate
    - 2-4 DLTs -> remain at the same dose

  - 5 DLTs -> de-escalate
  - 6-10 DLTs -> de-escalate and consider current dose as intolerable
- With 11 patients treated at current dose level
  - 0-1 DLT -> escalate
  - 2-5 DLTs -> remain at the same dose
  - 6-11 DLTs -> de-escalate and consider current dose as intolerable
- With 12 patients treated at current dose level
  - 0-2 DLTs -> escalate
  - 3-5 DLTs -> remain at the same dose
  - 6 DLTs -> de-escalate
  - 7-12 DLTs -> de-escalate and consider current dose as intolerable
- With 13 patients treated at current dose level
  - 0-2 DLTs -> escalate
  - 3-5 DLTs -> remain at the same dose
  - 6 DLTs -> de-escalate
  - 7-13 DLTs -> de-escalate and consider current dose as intolerable
- With 14 patients treated at current dose level
  - 0-2 DLTs -> escalate
  - 3-6 DLTs -> remain at the same dose
  - 7 DLTs -> de-escalate
  - 8-14 DLTs -> de-escalate and consider current dose as intolerable
- With 15 patients treated at current dose level
  - 0-2 DLTs -> escalate
  - 3-6 DLTs -> remain at the same dose
  - 7 DLTs -> de-escalate
  - 8-15 DLTs -> de-escalate and consider current dose as intolerable
- With 16 patients treated at current dose level
  - 0-2 DLTs -> escalate

  - 3-7 DLTs -> remain at the same dose
  - 8-16 DLTs -> de-escalate and consider current dose as intolerable

- With 17 patients treated at current dose level
  - 0-2 DLTs -> escalate
  - 3-7 DLTs -> remain at the same dose
  - 8 DLTs -> de-escalate
  - 9-17 DLTs -> de-escalate and consider current dose as intolerable

- With 18 patients treated at current dose level
  - 0-3 DLTs -> escalate
  - 4-7 DLTs -> remain at the same dose
  - 8 DLTs -> de-escalate
  - 9-18 DLTs -> de-escalate and consider current dose as intolerable