



## PROTOCOL A8081054

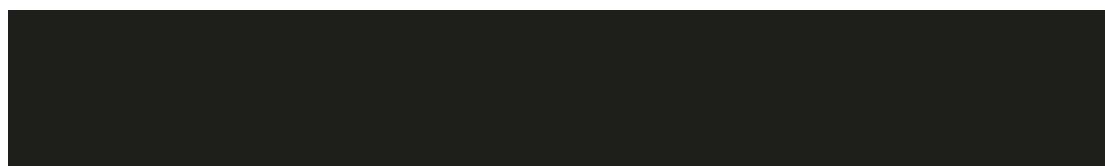
### A PHASE 1B STUDY OF CRIZOTINIB IN COMBINATION WITH PEMBROLIZUMAB (MK-3475) IN PATIENTS WITH UNTREATED ADVANCED ALK-TRANSLOCATED NON-SMALL CELL LUNG CANCER

## Statistical Analysis Plan (SAP)

**Version:** 1.0

**Author:** PPD

**Date:** June 26, 2015



**TABLE OF CONTENTS**

LIST OF TABLES .....	4
APPENDICES .....	4
1. AMENDMENTS FROM PREVIOUS VERSION(S) .....	5
2. INTRODUCTION .....	5
2.1. Study Design .....	5
2.2. Study Objectives .....	5
2.2.1. Primary Objective .....	5
2.2.2. Secondary Objectives .....	5
<b>CCI</b>	
3. INTERIM ANALYSES, FINAL ANALYSES, AND UNBLINDING .....	6
4. HYPOTHESES AND DECISION RULES .....	6
4.1. Statistical Hypotheses .....	6
4.2. Statistical Decision Rules .....	6
5. ANALYSIS SETS .....	7
5.1. Safety Analysis Set .....	7
5.2. Per Protocol Analysis Set (DLT-evaluable population) .....	7
5.3. Response Evaluable Analysis Set .....	7
5.4. PK Analysis Set .....	7
5.5. PRO Analysis Set .....	8
<b>CCI</b>	
5.7. Treatment Misallocations .....	8
5.8. Protocol Deviations .....	8
6. ENDPOINTS AND COVARIATES .....	8
6.1. Efficacy Endpoints .....	8
6.2. Safety Endpoints .....	11
6.2.1. Dose Limiting Toxicity (DLT) (Primary Endpoint for Dose Finding Cohorts) .....	11
6.2.2. Adverse Events .....	11
6.2.3. Laboratory Abnormalities .....	11
6.2.4. Other Safety Endpoints .....	11
6.3. PK Endpoints .....	11
<b>CCI</b>	

6.5. PRO Endpoints .....	12
6.6. Covariates .....	12
7. HANDLING OF MISSING VALUES .....	13
7.1. Missing Dates .....	13
7.2. Missing Efficacy Endpoint Values .....	13
7.3. Pharmacokinetics .....	13
<b>CCI</b>	
7.5. Patient Reported Outcomes .....	14
8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES .....	14
8.1. Statistical Methods .....	14
8.1.1. Statistical Methods for Dose De-Escalation: Up-and-Down Matrix Design with the mTPI Method .....	14
8.1.2. Methods for Estimating the MTD .....	16
8.1.3. Sample Size Determination .....	16
8.2. Statistical Methods for Different Types of Endpoints .....	17
8.3. Statistical Analyses .....	17
8.3.1. Standard Analyses .....	17
8.3.2. Analysis of Efficacy Endpoints .....	18
8.3.3. Analysis of Safety Endpoints .....	19
8.3.3.1. Analysis of Primary Endpoint .....	19
8.3.3.2. Analysis of Secondary Safety Endpoint .....	19
8.3.4. Analysis of Pharmacokinetics .....	23
8.3.4.1. Pharmacokinetic Analysis of Crizotinib and Pembrolizumab .....	23
8.3.4.2. Effect of Pembrolizumab on Crizotinib Pharmacokinetics .....	24
<b>CCI</b>	
<b>  </b>	
8.3.5. Analysis of Patient-Reported Outcomes .....	25
<b>CCI</b>	
<b>  </b>	
8.4. Summary of Secondary Efficacy Analyses .....	28
9. REFERENCES .....	29
10. APPENDICES .....	30

**LIST OF TABLES**

Table 1. Response Evaluation Criteria in Solid Tumors .....	38
---	----

**APPENDICES**

Appendix 1. Programming Specifications for AE Analyses .....	30
Appendix 2. Study Specific Information for Efficacy .....	34
Appendix 3. RECIST 1.1 .....	35
Appendix 4. Detailed Dose Re-Escalation/De-Escalation Scheme .....	39
Appendix 5. Visit Windows for the analysis of PRO endpoints.....	40

## 1. AMENDMENTS FROM PREVIOUS VERSION(S)

Not applicable.

## 2. INTRODUCTION

This document describes the planned statistical analyses for Protocol A8081054. 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 analysis plan will be described in the Clinical Study Report.

### 2.1. Study Design

*This is a Phase 1b, open-label, multicenter, multiple-dose, dose-finding, safety, and pharmacokinetic study of crizotinib in combination with pembrolizumab in sequential cohorts of adult patients with previously untreated ALK-positive advanced NSCLC with 2 phases (Dose Finding Phase and Dose Expansion Phase).*

*Approximately 70 total patients are expected to be enrolled in this study. The first phase of the study, the Dose Finding Phase, will utilize the mTPI method to determine the MTD of the combination regimen. The second phase, the Dose Expansion Phase, will further evaluate the combination regimen at the MTD to determine whether or not the MTD will also be the RP2D. The Dose Finding Phase of the trial will be completed when 10 DLT-evaluable patients have been treated at the highest DL associated with a DLT rate of <0.33. Crizotinib will be given orally (PO) either BID or QD, with or without food on a continuous daily dosing schedule. Pembrolizumab will be given as a 30-minute intravenous (IV) infusion Q3-weeks. Patients will continue with combination crizotinib and pembrolizumab treatment until RECIST-defined progression of disease and investigator decision that the patient is no longer experiencing clinical benefit, or unacceptable toxicity, death, or consent withdrawal, whichever occurs first.*

### 2.2. Study Objectives

#### 2.2.1. Primary Objective

- *To assess safety and tolerability of crizotinib in combination with pembrolizumab in the first-line treatment of patients with ALK-positive advanced non-squamous NSCLC in order to identify the Maximum Tolerated Dose (MTD) and select the Recommended Phase 2 Dose (RP2D).*

#### 2.2.2. Secondary Objectives

- *To evaluate the overall safety profile of crizotinib in combination with pembrolizumab;*
- *To document the anti-tumor activity of crizotinib in combination with pembrolizumab in previously untreated ALK-positive advanced NSCLC patients;*
- *To characterize the pharmacokinetics (PK) of crizotinib and pembrolizumab and to assess the effect of pembrolizumab on the PK of crizotinib;*

- *To evaluate tumor PD-L1 expression, as assessed by PD-L1 immunohistochemistry, as a predictor of anti-tumor activity.*



- *To evaluate the impact of combination therapy of crizotinib and pembrolizumab on patient-reported lung cancer symptoms (such as dyspnea, cough and pain) and global Quality of Life (QOL) per the validated cancer specific EORTC QLQ-C30, and its lung cancer module, QLQ-LC-13, and visual symptoms per the Visual Symptom Assessment Questionnaire (VSAQ-ALK).*

### 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 of efficacy will be performed approximately 18 months after enrollment of the last patient; 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 modified toxicity probability interval (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 de-escalation, no change in dose, or dose escalation.

##### Maximum Tolerated Dose Definition

The MTD estimate is the highest dose of crizotinib and pembrolizumab associated with the occurrence of DLTs in <33% of previously untreated ALK-positive advanced NSCLC patients.

### Dose Expansion Phase Cohort

Once the MTD (or the RP2D) for the combination has been defined, up to additional 40 patients with previously untreated ALK-positive advanced NSCLC will be enrolled and treated in the Expansion Phase (for a total of 50 patients treated at the MTD). The expansion phase will confirm the safety and tolerability as well as explore the anti-tumor activity of crizotinib in combination with pembrolizumab in patients with previously untreated ALK-positive advanced NSCLC.

## **5. ANALYSIS SETS**

### **5.1. Safety Analysis Set**

The safety analysis set includes all enrolled patients who receive at least 1 dose of crizotinib or pembrolizumab. This is the primary population for all standard analyses and is also the primary population for the efficacy endpoints of PFS and OS.

### **5.2. Per Protocol Analysis Set (DLT-evaluable population)**

The per protocol analysis set includes all patients enrolled in the Dose Finding Phase who are in the safety analysis population, and either experience DLT during the first 2 cycles, or complete the observation period for the first 2 cycles of treatment. Patients who do not receive at least 80% of the planned first 2 cycles of crizotinib dosing or both infusions of pembrolizumab within the DLT observation period due to reasons other than treatment-related AEs will not be evaluable for DLT.

### **5.3. Response Evaluable Analysis Set**

The response-evaluable population includes all patients in the safety analysis population who have an adequate baseline tumor assessment. This is the primary population for the efficacy endpoints of ORR, DR and TTR.

In addition, at any interim reporting of the data, patients also need to meet one of the following 2 criteria:

- have at least one post-baseline disease assessment
- withdraw from the trial or experience progression/death at any time on study

The post-baseline disease assessment must be at least 6 weeks from first dose (Cycle 1 Day 1)

### **5.4. PK Analysis Set**

The PK concentration population is defined as all patients in the safety analysis population who have at least 1 concentration of crizotinib, its metabolite PF-06260182, or pembrolizumab.

The PK parameter analysis population is defined as all patients in the safety analysis population who have at least 1 of the PK parameters of crizotinib, its metabolite PF-06260182, or pembrolizumab.

### 5.5. PRO Analysis Set

The PRO-evaluable population includes all patients in the safety analysis population who complete the patient reported questionnaires for at least one of the item both at baseline and at least 1 post baseline assessment.

CCI



### 5.7. Treatment Misallocations

Not applicable.

### 5.8. Protocol Deviations

The full list of protocol deviations will be compiled and presented in the study report. All deviations related to the statistical analyses or analyses populations will be described. As appropriate, sensitivity analyses may be conducted by excluding patients with specific protocol deviations.

## 6. ENDPOINTS AND COVARIATES

### 6.1. Efficacy Endpoints

Investigator assessment of tumor response will be made using RECIST version 1.1.<sup>3</sup>

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

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

Anti-tumor activity will be assessed by radiological tumor assessments conducted at baseline, then at Week 9, and every 6 weeks, thereafter. In addition, radiological tumor assessments will also be conducted to confirm a best response of CR or PR at  $\geq 4$  weeks after initial assessment of response, whenever disease progression is suspected (eg, symptomatic deterioration), and at the time of End of Treatment/Withdrawal (if not performed in the previous 6 weeks).

The primary efficacy summaries of tumor related data for this study will be based on a programmatic approach to derive tumor response/progression, using RECIST version 1.1.<sup>3</sup> As described in [Appendix 3](#) of this document, the derived tumor assessment will be based on the target lesion measurements, non-target lesion assessments, and new lesion records provided by the Investigator.

- **Progression Free Survival (PFS)** is defined as the time from the date of the first dose of crizotinib or pembrolizumab 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. PFS will be summarized in the safety analysis population

Patients with inadequate baseline assessments (adequate baseline is defined in [Appendix 2](#)) will have their event time censored on the date of first dose. Patients lacking an evaluation of tumor response after the date of the first dose or for whom the first on-study assessment of disease progression occurs after Week 17 (the first 2 tumor assessments + 2 weeks of allowance window), will also have their event time censored on the date of first dose unless death occurs within (and including) Week 17 (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, on study and progression free at the time of the analysis;
- Who discontinue treatment without documented disease progression and who do not die on study;
- Who have documentation of disease progression or death on study after  $\geq 2$ , consecutive missed tumor assessments (ie,  $> 14$  weeks after last on-study tumor assessment, 2 tumor assessments + 2 weeks of allowance window);
- Who are given anti-tumor treatment other than the study medication while on study and prior to documented disease progression or death on study. In this case, the last evaluable tumor assessment prior to start of the anti-tumor

treatment will be used. One exception to this rule is for patients who have documented PD or death within (and including) 14 days of anti-tumor treatment, in which case, the PD or death will be considered an event.

- **Eighteen-month PFS** is defined as the probability of being alive and progression-free at 18 months after the date of first dose based on the Kaplan Meier<sup>2</sup> estimate. The 18-month PFS rate is summarized as a product limit estimator based on the Kaplan-Meier method to account for censored events. Six-month and 12-month PFS are defined similarly.
- **Objective Response Rate (ORR)** is defined as the percent of patients with confirmed complete response (CR) or confirmed partial response (PR) according to RECIST v1.1, ([Appendix 3](#)) relative to the response evaluable population. 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.
- **Time to Response (TTR)** is defined as the time from first dose of crizotinib or pembrolizumab to the first documentation of objective tumor response (CR or PR) that is subsequently confirmed. TTR will only be calculated for the subgroup of patients with a confirmed objective tumor response.

$$\text{TTR (weeks)} = [\text{first date of Objective response (OR)} - \text{date of first dose} + 1]/7.02$$

- **Duration of Response (DR)** is defined as the time from first documentation of objective tumor response (CR or PR) that is subsequently confirmed, to the first documentation of objective tumor progression or to 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. Censoring for DR is identical to the censoring rules presented for PFS when patients have at least 1 on-study disease assessment. DR will only be calculated for the subgroup of patients with a confirmed objective tumor response.

$$\text{DR (weeks)} = [\text{progression/death date} - \text{first date of OR} + 1]/7.02$$

- **Overall Survival (OS)** is defined as the time from the first dose of crizotinib or pembrolizumab to the date of death due to any cause. OS will be summarized in the safety analysis population. For patients still alive at the time of the analysis, the OS time will be censored on the last date they were known to be alive. Patients lacking data beyond the day of first dose of study treatment will have their survival times censored at 1 day.

$$\text{OS (months)} = [\text{date of death} - \text{date of first dose} + 1]/30.44$$

- **Eighteen-month OS** is defined as the probability of being alive at 18 months after the date of first dose based on the Kaplan Meier estimate. The 18-month OS rate is summarized as a product limit estimator based on the Kaplan-Meier method to account for censored events. Twelve-months OS is defined similarly.

## 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 NCI CTCAE v 4.03. AEs meeting one of the definition criteria in Study Protocol Section 3.2 and occurring in the first 2 cycles of treatment (6 weeks) which are attributable to crizotinib, pembrolizumab or both will be classified as DLTs

### 6.2.2. Adverse Events

AEs as characterized by type, frequency, severity (as graded by NCI CTCAE version 4.03), timing, seriousness and relationship to study treatment.

#### Treatment Emergent Adverse Events

The focus of AE summaries will be on treatment-emergent AEs, those with initial onset or increasing in severity after the first dose of study treatment, through and including 28 calendar days after the last administration of crizotinib or 90 calendar days after the last administration of pembrolizumab, whichever is later, and before initiation of a new anti-cancer treatment.

#### Treatment Related Adverse Events

Treatment-related AEs are those judged by the Investigator to be at least possibly related to the study drugs (crizotinib and/or pembrolizumab) [with a cause related to study drug as indicated on the case report form (CRF)], or for which relatedness is recorded as “unknown” by the Investigator.

### 6.2.3. Laboratory Abnormalities

Laboratory abnormalities will be characterized by type, frequency, severity (as graded by NCI CTCAE v. 4.03) and timing. For laboratory tests without NCI CTCAE Grade definitions, results will be categorized as normal, abnormal, or not done.

### 6.2.4. Other Safety Endpoints

Descriptive statistics and categorical analyses will be presented for ECG results and vital signs data.

## 6.3. PK Endpoints

The following PK parameters of crizotinib, its metabolite PF-06260182, and pembrolizumab will be determined as data permits:

- Maximum plasma concentration ( $C_{max}$ ), time to maximum plasma concentration ( $T_{max}$ ), pre-dose concentration ( $C_{trough}$ ), area under the plasma concentration time curve from 0 to 8 hours ( $AUC_{0-8}$ ), area under the plasma concentration time curve over the dosing interval ( $AUC_{tau}$ ), and apparent plasma clearance ( $CL/F$ ) for crizotinib;
- $C_{max}$ ,  $T_{max}$ ,  $C_{trough}$ ,  $AUC_{0-8}$ ,  $AUC_{tau}$ , metabolite to parent ratio for  $AUC_{tau}$  ( $MRAUC_{tau}$ ), and metabolite to parent ratio for  $C_{max}$  ( $MRC_{max}$ ) for PF-06260182;
- $C_{trough}$  for pembrolizumab;
- Population-based PK parameters for crizotinib and pembrolizumab (if performed, will be presented in a separate report);
- CCI [REDACTED]

CCI [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

## 6.5. PRO Endpoints

- Overall change from baseline patient-reported lung cancer symptom, functioning, and global QOL scores using the EORTC QLQ-C30 and QLQ-LC-13 questionnaires;
- Frequency of occurrence and degree of bothersomeness of visual disturbances as reported by the patients on the VSAQ-ALK questionnaire.

## 6.6. Covariates

None.

## 7. HANDLING OF MISSING VALUES

### 7.1. Missing Dates

In compliance with Pfizer standards, imputation methods apply to partial dates. If the day of the month is missing for a start date used in a calculation, the 1<sup>st</sup> of the month will be used to replace the missing date. Similarly, if both the day and month are missing, the first day of the year is used. For stop dates, the last day of the month, or last day of the year is used if the day or both day and month are missing, respectively. These rules are used unless the calculations result in negative time durations (eg, date of resolution cannot be prior to date of onset). In these cases, the dates resulting in 0 time duration will be used. For PFS, TTR, DR and OS, if conventions result in a negative duration, durations will be reset to 1 day.

### 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 time to event endpoints, non-event observations will be censored and for ORR, patients with no post-baseline tumor evaluations will be counted as non-responders.

### 7.3. Pharmacokinetics

#### Concentrations below the limit of quantification:

For all calculations, figures and estimation of individual pharmacokinetic parameters, all concentrations assayed as below the level of quantification (BLQ) will be set to zero. In log-linear plots these values would not be represented. The BLQ values will be excluded from calculations of geometric means and their confidence intervals. A statement similar to “all values reported as BLQ have been replaced with zero” will be included as a footnote to the appropriate tables and figures.

#### 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 (ie, not done) or NS (ie, 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 patient's concentration data, the parameter will be coded as NC (ie, not calculated). (Note that NC values will not be generated beyond the day that a patient 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 (ie, analysis of variance), PK parameters coded as NC will also be set to missing.

If an individual patient has a known biased estimate of a PK parameter (due for example to an unexpected event such as vomiting before all of the study 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 study drug to be included in the concentration summary.

For evaluation of changes in PK of crizotinib when administered alone vs in combination with pembrolizumab, as assessed in patients who receive Dose Level -1 (DL-1), only patients with a matching pair of PK assessments under both conditions will be included in the PK summary. Patients who have been treated with 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.

CCI



## 7.5. Patient Reported Outcomes

For the EORTC QLQ-C30, QLQ-LC13 and VSAQ-ALK in cases where multiple answers are given to one item, the more severe answer (higher score for QLQ-C30, QLQ-LC13 and VSAQ-ALK) will be counted. If less than half of the constituent items have been answered for a multi-item subscale, that subscale will be considered missing. Single-item subscales will be considered missing if the constituent item is incomplete.

# 8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

## 8.1. Statistical Methods

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

The modified toxicity probability interval (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 ([Appendix 4](#)). Thus, compared to other advanced model-based designs published in the literature, the mTPI design is logically less complicated and easier to implement.

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 crizotinib and pembrolizumab,  $e_1$  is selected as 0.05, and  $e_2$  is selected as 0.03. Therefore, the target dosing interval for the DLT rate is  $(0.25, 0.33)$ .

The three dosing intervals are associated with three different dose-escalation decisions. The under-dosing interval corresponds to a dose re-escalation (RE), over-dosing corresponds to a dose de-escalation (D), and proper-dosing corresponds to staying at the current dose (S). Given a dosing interval and a probability distribution, the unit probability mass (UPM) of that dosing interval is defined as the probability of a patient 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).<sup>4</sup>

The Dose Finding Phase of the trial is completed when 10 DLT-evaluable patients have been treated at the highest dose associated with a DLT rate  $<0.33$ . For this reason, a modification has been applied to the original mTPI decision algorithm so that if 4 DLTs are observed at any dose level, that dose level is considered unacceptable and never used again in the remainder of the study since 4/10 patients with DLTs would be a  $\geq 33\%$  rate. In addition, a further modification will allow re-escalation and exposure of up to 10 patients at the higher dose in case of  $\leq 3$  out of 10 patients experience DLTs.

In case a de-escalation is required from the initial starting DL0, it is estimated that approximately up to 30 DLT-evaluable patients will need to be enrolled to estimate 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 10 DLT evaluable patients. We assume that higher doses of 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 study, the dose combination (crizotinib 200 mg BID, pembrolizumab 200 mg Q3weeks) may have a higher proportion of DLTs than, say, (crizotinib 250 mg BID, pembrolizumab 200 mg Q3weeks), and this variability may be simply related to small cohort size alone. To overcome this potential problem, we will 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.<sup>5</sup> 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 combinations to the Dose Expansion Phase, in case no clear choice exists between more than 1 competing MTD combination. While the limited sample size may result in up to 2 dose combinations of equal potential anti-tumor activity, under the circumstances of this study, likely only 1 will be chosen for the Dose Expansion Phase. This decision will be based upon the combination of data related to safety, available anti-tumor activity, and clinical judgment of the investigators and the Sponsor.

### 8.1.3. Sample Size Determination

The sample size 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 70 patients will be required to achieve all study objectives.

As far as the Dose Finding Phase of this study is concerned, 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 30 DLT evaluable patients will be enrolled in the dose escalation stage in order to have a reliable and accurate estimate of the MTD and determine the RP2D.

Subsequent patients will then enter the Dose Expansion Phase aimed at evaluating safety, anti-tumor activity and PK at the MTD.

The planned sample size for the Dose Expansion Phase is 50 patients (including 10 patients from the Dose Finding Phase of the study treated at the MTD) and was calculated using the 1-sample survival method. Assuming enrollment over 3 months and 18 months follow-up, 50 patients will allow to test the hypothesis that the true 18-month PFS probability is  $\geq 50\%$  vs.  $\leq 32\%$  for historical control (data from the crizotinib arm of Study A8081014)<sup>10</sup> with 85% power, alpha = 0.05 (1-sided), and based on exponential assumption.

## 8.2. Statistical Methods for Different Types of Endpoints

Listings and standard summary statistics will be used to summarize data from this study.

### Analysis of Time-to-Event Endpoints

Time-to-event endpoints will be summarized using the Kaplan-Meier method and displayed graphically when appropriate. Median event times (and other quartiles) and corresponding 2-sided 95% confidence interval (CI) will be provided (Brookmeyer R and Crowley JJ).<sup>1</sup>

### Analysis of Binary Endpoints

Binary endpoints will be summarized by percentages along with the 95% CIs using an exact method based on the F-distribution.

### Analysis of Continuous Endpoints

Continuous endpoints will be summarized by descriptive statistics, including the mean, standard deviation, median, minimum, and maximum values. CCI

### Analysis of Categorical Endpoints

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

## 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. 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, gender, height, weight, ethnicity, ECOG performance status, smoking history, prior anti-cancer therapy, histopathological classification, stage at diagnosis, extent of disease, 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 total number of cycles administered, the median (range) of cycles administered, dose intensity, dose modifications, dose interruptions, dose delays.

### 8.3.2. Analysis of Efficacy Endpoints

Efficacy analyses will be presented in the form of statistical summaries and data listings for the Dose Expansion Phase (including the 10 patients from the Dose Finding Phase of the study treated at the MTD). For the Dose Finding Phase, only data listings will be presented. For the Dose Expansion Phase, PFS and OS will be summarized in the safety analysis population; ORR will be summarized in the response-evaluable population; DR and TTR will be summarized for the subgroup of responding patients.

PFS, OS, and DR will be analyzed by the Kaplan-Meier method. The median event time (and other quartiles) and corresponding 2-sided 95% CI will be provided for each endpoint. Six, 12-, and 18-month PFS probabilities 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  $\log[-\log(6-, 12, 18\text{-month PFS probabilities})]$  will be calculated using normal approximation and then back transformed to give the CIs for the respective PFS probabilities. Twelve- and 18-month OS probabilities will be calculated similarly.

TTR will be summarized with descriptive statistics. Best overall response (BOR) will be summarized and ORR will be calculated along with the corresponding exact 2-sided 95% CI using the exact method based on the F-distribution

#### Analysis Sets

- Response-evaluable population for ORR (Dose Expansion Phase including the 10 patients from the Dose Finding Phase of the study treated at the MTD).
- Safety analysis population for PFS and OS (Dose Expansion Phase including the 10 patients from the Dose Finding Phase of the study treated at the MTD).
- Patients with an overall objective response of CR or PR in the response-evaluable population for DR and TTR (Dose Expansion Phase including the 10 patients from the Dose Finding Phase of the study treated at the MTD).

Listings or/and tables (when applicable) will be sorted by dose level.

#### Tumor Response

- The best overall response is the best response (CR, PR, SD, PD, Early Death or IND) recorded during the “on-study” period (as defined in [Appendix 2](#) of this document).
- Best overall response is determined from the sequence of objective responses.
- A patient’s best overall response will be Early Death if the patient died within 42 days of treatments start and prior to having sufficient evaluations for overall response.

For BOR, if a patient has not achieved an objective response, but remained stable for at least 6 weeks after starting treatment, then the BOR for such a patient will be SD.

### **8.3.3. Analysis of Safety Endpoints**

#### **Analysis Set**

- Summaries and analyses of the primary safety endpoint will be based on the DLT-evaluable analysis population. All other summaries and analyses of safety parameters will be based on the safety analysis population. Safety data will be summarized by dose level and/or by dose level within dosing regimen using appropriate tabulations and descriptive statistics (pooling together patients treated at the MTD from both phases of the study). If applicable, safety data collected during the crizotinib lead-in period for patients who are treated at DL-1 will be reported separately.

#### **8.3.3.1. Analysis of Primary Endpoint**

Dose Limiting Toxicity (DLT) is the primary endpoint of the study. The occurrence of DLTs observed in the Dose Finding Phase will be used to estimate the MTD as described in Study Protocol Section 3.2. Adverse events constituting DLTs will be listed per dose level.

#### **DLT-related listings will be produced by dose level and will include:**

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

#### **8.3.3.2. Analysis of Secondary Safety Endpoint**

#### **Adverse Events**

Adverse Events (AEs) will be graded by the investigator according to the 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, those with initial onset or increasing in severity after the first dose of study treatment.

Because the frequency of certain medical concepts or conditions may be underestimated by reliance on single MedDRA preferred terms, certain preferred terms will be analyzed in aggregate using clustered terms. Patients having more than 1 AE preferred term within a clustered term will contribute 1 event to the clustered term at the highest grade observed. Clustered terms to be used for summary of data from this study will be defined prior to data summary based on information included in the current Safety and Risk Plan maintained by the Sponsor.

The number and percentage of patients who experienced any treatment-emergent AE, treatment-emergent SAE, treatment-related AE, and treatment-related SAE will be summarized according to worst toxicity grades. The summaries will present AEs both on the entire study period and by cycle (Cycle 1, Cycle 2, and Cycles  $\geq 3$ ). In DL-1, the summaries will present AEs during the lead-in period, overall, and by cycle (Cycle 1, Cycle 2, and Cycles  $\geq 3$ ).

- 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 with permanent/temporary discontinuation or dose reduction, if applicable, because of an AE will be presented.
- All treatment-emergent AEs (all causality) will be summarized by MedDRA preferred term (PT) and/or clusters of MedDRA PT. A summary of AEs by preferred term/clustered term and maximum CTCAE grade will be presented, in decreasing order of frequency. A summary of AEs by preferred term/clustered term and maximum CTCAE grade group (Grade 1-2, Grade 3-4, and Grade 5) will also be presented, in decreasing order of frequency. Treatment-emergent AEs will be summarized by dose level.
- Treatment-related AEs, all-causality SAEs and treatment-related SAEs will be summarized similarly to all causality treatment-emergent AEs. Displays of treatment related events will include events judged by the Investigator to be at least possibly related/unknown to one of the study drugs (crizotinib and/or pembrolizumab). Displays will not be focused on crizotinib-related only events or pembrolizumab-related only events.
- Treatment-emergent AEs associated with permanent/temporary discontinuation of the study drug or with dose reduction, if applicable, will be summarized (taking into consideration the action taken from the CRF AE page).
- Deaths will be summarized by two time periods: 1: within 28 days of last dose of crizotinib and/or within 90 days of last dose of pembrolizumab and 2:  $>28$  calendar days after the last administration of crizotinib and  $>90$  calendar days after the last administration of pembrolizumab, and cause of death. Deaths that occurred within 28 days after the last dose of crizotinib and/or within 90 days of last dose of pembrolizumab are defined as on-treatment deaths. Death data will also be listed.

Additional analyses may be performed for AEs of special interest: Time to AE onset (in days), Duration of AE (in days) and Prevalence of AE (definitions for time to AEs, duration of AEs and prevalence of AEs are provided in [Appendix 1](#) of this document).

Duration of AE (days) will be summarized using the Kaplan-Meier method. The median and other quartiles, as appropriate, will be presented along with associated 95% CIs. Descriptive statistics will be presented for time to AE onset (days) and duration of AEs for the subgroup of subjects with the AE, if the number of events is small for reliable use of the Kaplan-Meier method. The prevalence of AEs of special interest will be summarized by maximum CTC severity grade. Programming specifications for AE analyses are described in [Appendix 1](#).

### **Laboratory Tests**

The number and percentage of patients who experience laboratory test abnormalities will be summarized according to the worst toxicity grade observed for each laboratory assay. The analyses will summarize laboratory tests during the lead-in period (only applicable to DL-1), overall, and by cycle (Cycle 1, Cycle 2, and Cycles  $\geq 3$ ). Shift tables will be provided to examine the distribution of laboratory abnormalities.

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

- **Hematology and blood chemistry** results will be graded according to the NCI CTCAE Version 4.03. A shift summary of baseline grade by maximum post-baseline CTCAE grade will be presented, as appropriate. Shift tables may also be summarized separately for Cycle 1, Cycle 2, and  $>$  Cycle 2. Patients who developed toxicities of grade  $\geq 3$  will be listed.
- **E-DISH** scatter plots of maximum ALT and maximum AST vs. maximum total bilirubin on-study will be presented
- **Other Laboratory Tests** – Individual patient test results will be listed.

### **ECG**

The analysis of ECG results will be based on data collected during the first 3 cycles for all patients except in DL-1 where ECG collection will also be performed during the crizotinib lead-in period. In this case baseline assessment will be the one obtained prior to start of crizotinib.

ECG measurements (an average of the triplicate measurements) will be used for all data presentations. Any data obtained from ECGs repeated for safety reasons after the nominal time points will not be averaged along with the preceding triplicates. Interval measurements from repeated ECGs will be included in the outlier analysis (categorical analysis) as individual values obtained at unscheduled time points.

QT intervals will be corrected for heart rate (QTc) using standard correction factors (ie, Fridericia's [QTcF; default correction], or Bazett's [QTcB]). Data will be summarized and listed for HR, RR interval, PR interval, QRS complex, QTcF and QTcB, and by dose level. Categorical analysis of the ECG parameters will be performed as follows:

- The number and percentage of patients with maximum increase from baseline in QTcF/QTcB (<30, 30- 60, and  $\geq$ 60 ms);
- The number of and percentage patients with maximum post-dose QTcF/QTcB (<450, 450- $\leq$ 480, 480-  $\leq$ 500, and  $\geq$ 500 ms);
- PR interval changes from baseline  $\geq$ 50% if absolute baseline value was <200 ms, and  $\geq$ 25% if absolute baseline value was  $\geq$ 200 ms;
- QRS complex changes from baseline  $\geq$ 50% if absolute baseline value was <100 ms, and  $\geq$ 25% if absolute baseline value was  $\geq$ 100 ms;

Shift tables will be provided for baseline vs. worst on study QTcF and QTcB using maximum CTCAE Grade.

### **Vital Signs**

The number and percent of patients in each of the following minimum and maximum blood pressure, body weight and pulse rate categories will be presented:

Vital Sign	Category
Blood Pressure	Maximum Change from baseline (increase or decrease) in SBP of $\geq$ 40 mmHg
	Maximum Change from baseline (decrease) in SBP of $\leq$ 60 mmHg
	Maximum Change from baseline in DBP (increase or decrease) of $\leq$ 20 mmHg
	Maximum Change from baseline in DBP (decrease) of $\leq$ 40 mmHg
Body Weight	Maximum change from baseline body weight (increase or decrease) $\geq$ 10%
Pulse Rate	Minimum Pulse Rate <50 bpm
	Maximum Pulse Rate > 120 bpm
	Maximum Change from baseline (increase or decrease) $\leq$ 30 bpm

In addition, the baseline and the change from baseline in blood pressure and pulse rate will be summarized using descriptive statistics by visit. For DL-1 baseline assessment will be the one obtained prior to start of crizotinib.

### **ECOG Performance Status**

ECOG performance status data will be summarized with a shift table of the baseline and the worst on study status.

## **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 study start are considered previous medications. All ongoing medications at study start or with start date after study start 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 patients 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, prior and follow-up systemic therapy for the primary diagnosis will be listed for each patient and summarized as appropriate.

## **Data Safety Monitoring Committee**

A Data Safety Monitoring Committee will not be established for the study. However, Pfizer procedures for periodic safety review will be applied, as described in Section 9.7 of the study protocol.

### **8.3.4. Analysis of Pharmacokinetics**

Analysis Set: PK

#### **8.3.4.1. Pharmacokinetic Analysis of Crizotinib and Pembrolizumab**

Descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% confidence interval) for all plasma/serum concentrations of crizotinib (plasma) and pembrolizumab (serum) will be presented in tabular form by dose level, visit (cycle/day), and nominal time.

When applicable, plots of mean and median crizotinib and pembrolizumab plasma/serum concentrations by nominal time will be presented by dose level, visit (cycle/day) on both linear-linear and log-linear scales. Similarly, individual patient profiles of concentration-time data will be plotted by treatment, dose level, cycle, and day, using nominal times.

Steady state trough plasma concentration ( $C_{trough,ss}$ ) of crizotinib and its metabolite PF-06260182 is defined as predose plasma concentration following at least 14 consecutive days of the prescribed dose without dosing interruption. Individual mean steady state predose concentration ( $C_{trough, ss, mean}$ ) of crizotinib and metabolite PF-06260182 will be obtained by using the arithmetic mean of all  $C_{trough, ss}$  values for each patient. Presentations for  $C_{trough,ss}$  and  $C_{trough, ss, mean}$  of crizotinib and its metabolite PF-06260182 were to include, but not limited to:

- Summary of  $C_{trough,ss}$  by visit (Cycle/Day) and by ethnicity and race group (Asians versus Non-Asians) using descriptive statistics;

- Summary of  $C_{trough,ss,mean}$  by ethnicity and race group (Asians versus Non-Asians) using descriptive statistics;
- Linear plots of median/mean  $C_{trough,ss}$  against visit (Cycle/Day) by ethnicity and race group (Asians versus Non-Asians).

For DL-1, standard plasma PK parameters for crizotinib and its metabolite PF-06260182 will be estimated using non-compartmental methods for PK concentration data collected on Cycle -1 Day 15 (when crizotinib given alone) and Cycle 6 Day 1 (in combination with pembrolizumab). PK parameters for crizotinib will include:  $C_{max}$ ,  $T_{max}$ ,  $C_{trough}$ ,  $AUC_{0-8}$ ,  $AUC_{tau}$ , and oral plasma clearance (CL/F). PK parameters for PF-06260182 will include:  $C_{max}$ ,  $T_{max}$ ,  $C_{trough}$ ,  $AUC_{0-8}$ ,  $AUC_{tau}$ ,  $MRAUC_{tau}$ , and  $MRC_{max}$ . PK parameters will be listed and summarized in tabular form by visit (Cycle/Day) or treatment (when given alone and in combination), using descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% confidence interval). For  $T_{max}$ , the range (min, max) will also be provided. All calculations will follow the Pfizer Clinical Pharmacology Guidance “Pharmacokinetic Data Handling and Non-Compartmental Analysis Conventions”. Box plots for  $AUC_{tau}$  and  $C_{max}$  for crizotinib and PF-06260182 (when given alone and in combination with pembrolizumab) will be generated, respectively. Individual data points, the geometric mean and the median of the parameter in each treatment will be overlaid on the box plots.

Trough concentrations for pembrolizumab will be plotted for each dose using a box-whisker or linear plots by visit (cycle/day) in order to assess the attainment of steady-state.

In addition, Non-linear Mixed Effects Modeling (NONMEM) approaches will be explored to further describe the PK profile of pembrolizumab in combination with crizotinib.

Additional tables, listings, and figures for post-hoc analyses may be included in the final CSR.

#### **8.3.4.2. Effect of Pembrolizumab on Crizotinib Pharmacokinetics**

For DL-1, the effect of pembrolizumab on crizotinib PK will be evaluated using steady-state crizotinib  $AUC_{tau}$  and  $C_{max}$  as the primary pharmacokinetic parameters after crizotinib is administered as a single agent (Cycle -1 Day 15 [Reference]) and in combination with pembrolizumab (Cycle 6 Day 1 [Test]). The  $AUC_{tau}$  and  $C_{max}$  will be log transformed and analyzed using a mixed effect model with treatment as the fixed effect and patient as the random effect. The ninety percent confidence interval for the ratio of geometric means of  $AUC_{tau}$  and  $C_{max}$  (Test/Reference) will be computed to assess the effect magnitude. The effect of pembrolizumab on the PK of PF-06260182 may also be assessed using this method.

For the other dose levels that do not incorporate a crizotinib lead-in period, the effect of repeated pembrolizumab doses on crizotinib PK will be evaluated by comparing the steady-state crizotinib and PF-06260182 PK to that observed in prior single-agent crizotinib studies, as appropriate.

CCI



### 8.3.5. Analysis of Patient-Reported Outcomes

The PRO-evaluable population will be the primary population for the analysis of PROs. Change from baseline scores over treatment will be evaluated in the PRO evaluable population. Visit windows will be applied for the analysis of the PRO endpoints. Further details are described in [Appendix 5](#).

#### PRO Scoring Procedure

The subscales of the EORTC QLQ-C30 and the QLQ-LC13 will be scored based on the EORTC scoring manual. In summary, each scale of the EORTC QLQ-C30 and the QLQ-LC13 will be transformed so that scale scores will range from 0 to 100. The transformation will proceed in two steps. First, the average of the items contributing to a subscale will be calculated to compute the raw score of the scale. Next, a linear transformation will be applied to ‘standardize’ the raw score. After scores are transformed, higher scores on the EORTC QLQ-C30 or the QLQ-LC13 will represent higher (“better”) levels of functioning and/or a higher (“worse”) level of symptoms.<sup>6,7,8</sup>

For the VSAQ-ALK questionnaires descriptive statistics will be used to evaluate each item. Item 7 was developed based on question 6 (impairment on daily activities) of the Work Productivity Questionnaire (WPAI SHP). Item 7 will be scored according to the WPAI SHP guideline for question 6.

#### Instrument Compliance Rates

At each time point, the number and percentage of patients who completed the QLQ-C30, QLQ-LC13 and VSAQ-ALK will be summarized. A questionnaire is considered complete if at least 1 item was answered by the patient. Scoring and handling of missing data for the EORTC QLQ-C30 and QLQ-LC-13 will be implemented in line with the scoring manual.

**Global QOL, Functioning and Lung Cancer Symptoms.**

Summary statistics (mean, standard deviation, median, range, and 95% CI) of actual scores and change from baseline at each time point will be calculated for all scales of the EORTC QLQ-C30 and QLQ-LC13 scores.

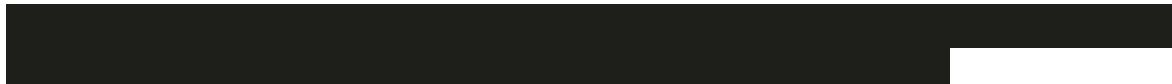
The number and proportion of patients who improved, worsened, or remained stable as compared to baseline will also be summarized for all of the symptom and functional domains, global QOL, and single items of the EORTC QLQ-C30 and QLQ-LC-13. Patients will be classified as improved, stable or deteriorated according to a 10-point or greater change in score from baseline as this is perceived by patients as being clinically significant on the EORTC QLQ-C30 (Osoba).<sup>9</sup>

If less than half of the constituent items on the QLQ-C30 and QLQ-LC13 have been answered for a multi-item subscale, that subscale will be considered missing. Single-item subscales will be considered missing if the constituent item is incomplete.

**VSAQ-ALK (Visual Symptoms)**

Frequency analyses will be performed and reported for each item of the VSAQ. For the VSAQ-ALK questions not answered will be considered missing items and will not be utilized.

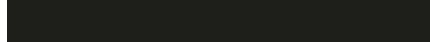
CCI



CCI



CCI



#### 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) ( <a href="#">Section 8.3.2</a> )	See <a href="#">Section 6.1</a>	Secondary analysis
DR	Subgroup of pts with OR from the Response Evaluable Population	K-M method (median and 95% CI) ( <a href="#">Section 8.3.2</a> )	See <a href="#">Section 6.1</a>	Secondary analysis
TTR	Subgroup of pts with OR from the Response Evaluable Population	Univariate (median and range) ( <a href="#">Section 8.3.2</a> )	See <a href="#">Section 6.1</a>	Secondary analysis
PFS	Safety Analysis	K-M method (median, 6-, 12-, 18-month PFS and 95% CIs) ( <a href="#">Section 8.3.2</a> )	Censor patients on the day of the last evaluable tumor assessment documenting absence of disease progression for ... ( <a href="#">See Section 6.1</a> )	Secondary analysis
OS	Safety Analysis	K-M method (median, 12- and 18-month OS and 95% CIs) ( <a href="#">Section 8.3.2</a> )	Censor patients on the day last known to be alive ( <a href="#">See Section 6.1</a> )	Secondary analysis

CI: Confidence intervals; DR: Duration of Response; K-M: Kaplan-Meier; OR: Objective Response; ORR: Objective Response Rate; PFS: Progression-free Survival; TTR: Time to Response; OS: Overall Survival

## 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. Ji Y et al. A modified toxicity probability interval method for dose-finding trials. *Clinical Trials* 2010; 7:653 663.
5. Dykstra R, Robertson T. An algorithm for isotonic regression for two or more independent variables. *Ann Stat*. 1982;10:708-716.
6. Aaronson NK 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*. 85:365-76, 1993.
7. Aaronson NK et al. The European Organization for Research and Treatment of Cancer (EORTC) modular approach to quality of life assessment in oncology: an update, in *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2nd Ed. S. B, Editor. Raven Press: New York. 179-89, 1996.
8. Bergman B et al. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *Eur J Cancer*. 30A:635-42, 1994.
9. Osoba D et al, Interpreting the significance of changes in health related quality of life scores. *J. Clin Oncology*. 1998. 16:139-144.
10. Mok T, Kim D-W, Wu Y-L, et al. First-line crizotinib versus pemetrexed-cisplatin or pemetrexed-carboplatin in patients (pts) with advanced ALK-positive non-squamous non-small cell lung cancer (PROFILE 1014). *J Clin Oncol*. 2014 ; 32 : 5s (suppl ; abstr. 8002).

## 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  $\text{first AE start date} - \text{first dose date} + 1$ .

*AE start date*

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

*First dose date*

The date of the starting dose: The date of first dose of study medication based on the Dosing CRF page.

#### 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  $\text{AE-end date} - \text{AE start date} + 1$  excluding any overlap. Duration of AE is defined for patients 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 patient 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 study visit date. Patients 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 patient will be censored at that date (ie, 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*

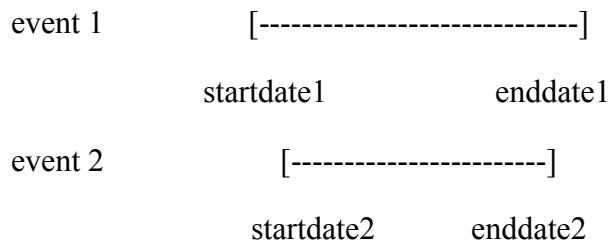
Death date is based on the Notice of Death CRF page.

### 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 (ie, 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.

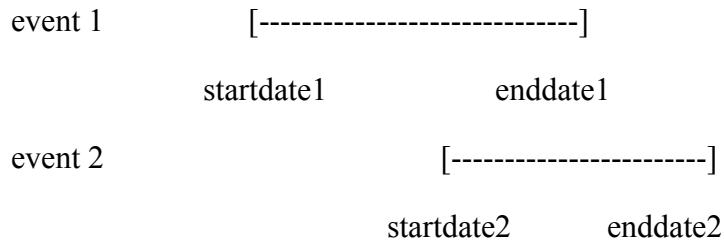
The following scenarios provide examples of the calculation for 2 events in the same cluster. The extension to 3 or more events of the same cluster is similar.

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



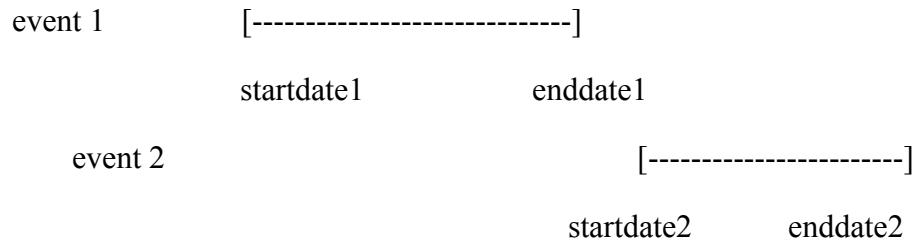
$$duration = enddate1 - startdate1 + 1$$

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



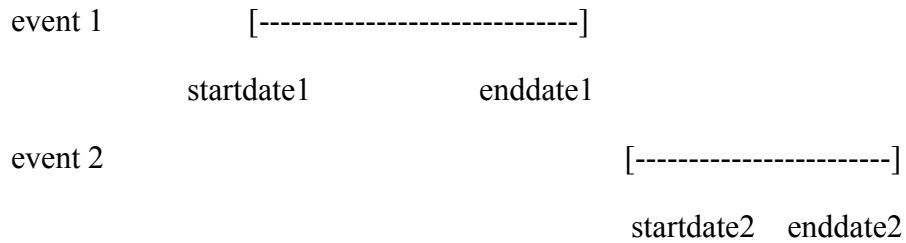
$$duration = enddate2 - startdate1 + 1$$

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



$$duration = enddate2 - startdate1 + 1$$

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



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

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 patients except those who either have discontinued or died prior to the specified time period.

AE prevalence may be presented separately for Cycle 1, Cycle 2, Cycle 3, Cycle 4, Cycle 5, Cycle 6, and Cycle  $>6$ , and Lead-in period, for DL-1.

2. **Assumptions**

- 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 DL-1 and for lead-in 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 period, for DL-1, that is still ongoing in Cycle 1 (ie, did not have a resolution date in the Lead-in phase).
- For Cycle 2, a patient is counted in the numerator if the patient has an onset date of the AE during Cycle 2 OR an onset date during the Lead-in period, for DL-1, or Cycle 1 that is still ongoing in Cycle 2 (ie, did not have a resolution date in the Lead-in phase or Cycle 1). The calculation for Cycles 3, 4, 5, and 6 is similar.
- For Cycles  $>6$ , a patient is counted in the numerator if the patient has an onset date in Cycles  $>6$  OR an onset date in an earlier cycle that is still ongoing (ie, did not have a resolution date in Cycle 6 or earlier cycle).

- The denominator for a particular time period will include patients who are at risk prior to the time period. The number at risk includes all patients except those who either have discontinued both study drugs (based on the 'Subject Summary End of Treatment Crizotinib' and on the 'Subject Summary End of Treatment Pembrolizumab' CRF) or died prior to the specified time period (ie, 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 study treatment
- **Adequate Baseline:**
  - Baseline tumor evaluations must be performed within 4 weeks (28 days) prior to the first dose of study treatment; tumor scans performed outside this window will be reviewed and might be considered acceptable on a case by case basis.;
  - Presence of at least one measurable lesion per RECIST version 1.1;
  - 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, MRI, Physical Exam, Bone Scan and Other. Note: If based on data review “unacceptable” methods (eg, 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 until patient death, progression of disease, patient no longer willing to participate, start of other anti-cancer treatment, or 28 days after last administration of crizotinib (or 90 days after last administration of pembrolizumab, whichever is later), whichever is earlier.
- **Subsequent anti-tumor treatment:** includes any systemic anticancer therapy (other than study medication), radiation and surgery for removal of target or non-target lesions (resected or partially resected).

### Appendix 3. RECIST 1.1

The determination of anti-tumor efficacy during this study will be based on objective tumor assessments made according to the RECIST version 1.1 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  $\geq 15$  mm 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  $< 10$  mm or pathological lymph nodes with  $\geq 10$  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 after baseline. 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 non-nodal target lesions and the short axis dimension will be used for each target lymph node. The sum of the longest diameter for target lesions/short axis dimension for target lymph nodes will be calculated and recorded as the baseline sum of lesion dimensions (SLD) 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 anti-tumor effect of a treatment.

### **Definitions of Tumor Response**

#### **Target Lesions**

**Complete Response (CR)** is defined by the disappearance of all non-lymph node target lesions (where all target lesions are recorded with a length of 0 mm on the “Target Lesions” eCRF). Any pathological lymph nodes (recorded as target lesion) must have reduction in short axis to <10 mm. Note: the SLD may not be zero if lymph nodes are included as target lesions.

**Partial Response (PR)** is defined by a 30% or more decrease in SLD of target lesions, taking as reference the baseline SLD.

**Progressive Disease (PD)** is defined by a 20% or more increase in the SLD of target lesions relative to baseline or the smallest SLD (nadir) recorded since first dose. In addition to the relative increase of 20%, SLD must also demonstrate an absolute increase of at least 5 mm ( $\geq 5$  mm) relative to baseline or the smallest SLD (nadir) recorded since first dose.

**Stable Disease (SD)** is assigned when neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD is observed, taking as reference the smallest sum diameters while on study.

**Indeterminate (IND)** is assigned if any individual target lesion is evaluated as “Indeterminate” or if inconsistent methods are used for post-baseline lesion assessment for any lesion, or if one or more target lesions are not assessed.

#### **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 (<10 mm short axis).

**Non-CR/Non-PD** is defined as a persistence of  $\geq 1$  non-target lesions.

**Progressive disease (PD)** is defined as unequivocal progression of existing non-target lesions, or the appearance of  $\geq 1$  new lesion. PD is assigned if any non-target lesion is marked “Increased” on the “Non-Target Lesion” eCRF. However, in an effort to programmatically define “unequivocal progression” of non-target lesions, the derived non-target lesion response will also take into account the “Non-Target Lesions” assessment from the IOTA eCRF.

**Indeterminate (IND)** is assigned if any individual non-target lesion is evaluated as “Indeterminate” or if inconsistent methods are used for post-baseline lesion assessment for any lesion, or if one or more non-target lesions are not assessed

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  $\geq 4$  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 6 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.

**Table 1. Response Evaluation Criteria in Solid Tumors**

Target Lesion Response	Non-Target Lesion Response	Overall Response
CR	CR	CR
CR	Non-CR/Non-PD	PR
CR	PD	PD
CR	IND	PR
PR	CR	PR
PR	Non-CR/Non-PD	PR
PR	PD	PD
PR	IND	PR
SD	CR	SD
SD	Non-CR/Non-PD	SD
SD	PD	PD
SD	IND	SD
PD	Any Response	PD
IND	PD	PD
IND	Non-PD	IND
CR/PR/SD/PD/IND	Not Collected at Baseline	CR/PR/SD/PD/IND

CR=Complete Response, PR=Partial Response, SD=Stable Disease, PD=Progressive Disease, IND=Indeterminate.

Note: If non-target (or target) lesions are not collected at baseline, then the overall response is equivalent to the target (or non-target) lesions response.

**Best Overall Response Evaluation for Each Patient:**

- The best overall response is the best response (CR, PR, SD, Early Death, PD or IND) recorded during the “on-study” period for each patient.
- Best overall response is derived from the sequence of objective responses.
- A patient’s best overall response will be Early Death if the patient died within 42 days of date of the first dose and prior to having sufficient evaluations for overall response.
- Assessments done after PD or after “anti-tumor treatment” but prior to PD will not be considered for evaluation of best overall response.
- For a patient to qualify for a best response of SD, the overall response evaluation must have met the stable disease criteria at least once since the date of the first dose at a minimum interval of 6 weeks (42 days).
- Unconfirmed CR and PR will be classified as a best response of SD provided it meets the 42 days requirement.
- Indeterminate (IND) is assigned for a patient who has only a baseline assessment, or a response assessment of CR/PR/SD at an interval less than 6 weeks and has no subsequent disease evaluation, or has all overall response evaluations assessed as IND.

#### Appendix 4. Detailed Dose Re-Escalation/De-Escalation Scheme

		Number of Patients Treated at Current Dose									
Number of tDLTs	1	2	3	4	5	6	7	8	9	10	
	0	NA	NA	RE	RE	RE	RE	RE	RE	RE	
	1	NA	NA	S	S	RE	RE	RE	RE	RE	
	2		NA	D	S	S	S	S	RE**	RE**	
	3			DU	DU	D	S	S	S	RE**	
	4*				DU	DU	DU	DU	DU	DU	
	5					DU	DU	DU	DU	DU	
	6						DU	DU	DU	DU	
	7							DU	DU	DU	
	8								DU	DU	
	9									DU	
	10									DU	

\*A modification has been applied to the original mTPI algorithm. Since this protocol specifies a DLT rate <33% in 10 patients, all entries on the row with 4 toxicities are fixed to DU since 4/10 would be ≥33%.

\*\* This further modification will allow re-escalation and exposure of up to 10 patients at the higher dose if ≤ 3 out of 10 patients with a DLT at the current dose and assuming the previous higher dose was not already a DU.

RE = Re-escalate to the next higher dose or if current dose level is DL0 stay on DL0.

S = Stay at the current dose.

D = De-escalate to the next lower dose.

U = The current dose is unacceptably toxic.

NA = Not applicable. The first three patients will have to be evaluable for DLT before assigning DL for the next patient.

Targeted DLT rate at MTD <33%.

## Appendix 5. Visit Windows for the Analysis of PRO Endpoints

The following visit label and visit windows will be applied for the analysis of PRO endpoints.

A Questionnaire is considered complete if at least one question is answered regardless of whether DONE/ NOT DONE is checked on the case report form.

In the case of multiple records for a patient within a particular visit window, then the average of the response values will be taken.

Visit Label	Visit Window (inclusive)	Target Day	Width (days)
Cycle 1 Day 1	<= 1		
Cycle 2	12 to 32	22	21
Cycle 3	33 to 53	43	21
Cycle 4	54 to 74	64	21
Cycle 5	75 to 95	85	21
Cycle 6	96 to 116	106	21
Cycle 7	117 to 137	127	21
Cycle 8	138 to 158	148	21
Cycle 9	159 to 179	169	21
Cycle 10	180 to 200	190	21
Cycle 11	201 to 221	211	21
Cycle 12	222 to 242	232	21
Cycle 13	243 to 263	253	21
Cycle 14	264 to 284	274	21
Cycle 15	285 to 305	295	21
Cycle 16	306 to 326	316	21
Cycle 17	327 to 347	337	21
Cycle 18	348 to 368	358	21
Cycle 19	369 to 389	379	21
Cycle 20	390 to 410	400	21
.....			
.....			
End of Treatment	End of Patient's Treatment		