Title: A Phase 3 Multicenter Open-label Study of Brigatinib (AP26113) versus Crizotinib in Patients with ALK-positive Advanced Lung Cancer

NCT Number: NCT02737501

SAP Approve Date: 2020-06-03

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

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

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

STATISTICAL ANALYSIS PLAN

STUDY TITLE: A Phase 3 Multicenter Open-label Study of Brigatinib (AP26113) versus Crizotinib in Patients with ALK-positive Advanced Lung Cancer

PROTOCOL NUMBER:	AP26113-13-301
PROTOCOL DATE:	22 October 2015
PROTOCOL VERSION:	Version 4.0
PROTOCOL AMENDMENT DATE:	2020-05-12
SAP VERSION:	Version 5.0
DATE:	2020-06-03

TABLE OF CONTENTS

TABLE OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Term
AE	adverse event
ALK	anaplastic lymphoma kinase
ALT	alanine aminotransferase
ARIAD	ARIAD Pharmaceuticals, Inc.
AST	aspartate aminotransferase
BIRC	blinded independent review committee
CNS	central nervous system (see iCNS)
CR	complete response
CSR	Clinical Study Report
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organisation for Research and Treatment of Cancer
HRQoL	health-related quality-of-life
iCNS	intracranial Central Nervous System
ITT	Intention-to-treat
MRI	magnetic resonance imaging
NE	not evaluable
NSCLC	non-small cell lung cancer
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PK	pharmacokinetic(s)
PR	partial response
PRA	PRA Health Sciences
PRO	patient-reported outcomes
QLQ	Quality of Life Questionnaire
QTcF	QT interval corrected (Fridericia)
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SRS	stereotactic radiosurgery
TKI	tyrosine kinase inhibitor
ULN	upper limit of normal

1 INTRODUCTION

This statistical analysis plan (SAP) describes the study design and statistical analyses for study protocol AP26113-13-301, entitled "A Phase 3 Multicenter Open-label Study of Brigatinib (AP26113) versus Crizotinib in Patients with ALK-positive Advanced Lung Cancer" by ARIAD Pharmaceuticals, Inc.

This statistical analysis plan (SAP) has been prepared to support a clinical study report (CSR) that will be included in regulatory submissions intended to achieve marketing approval. CSR analyses, as directed by this document, may also be used for other purposes such as publication of study results, and analyses will be rerun on updated data. Evaluation of the primary endpoint will be assessed in up to two interim analyses and one final analysis based on numbers of progression-free survival (PFS) events observed as described in as described in Section 3.4.3.5. Since study AP26113-13-301 will be ongoing at the time these evaluations take place additional analyses as specified in this SAP are anticipated at least annually.

Changes from analyses planned in the study protocol will be accounted for in this document. Analysis specifications in this document supersede suggestions for of descriptions of analyses contained in other documents, such as specifications prepared by external vendors, associated with study AP26113-13-301.

For revisions to the CSR any changes and additions to analyses specified here will be detailed in a companion document that will be included as an appendix.

An administrative update (Version 2) was made to this plan on February 18, 2018, consisting of clarifications of text that were requested when creating programming specifications and analysis programs based on this document.

A further update (Version 3) was made to this plan on March 27, 2018, in which substantive changes were made.

- A statement was added in Section 3.4.3.6 that formal statistical inference will be completed at the first successful evaluation of the primary endpoint.
- Section 3.5.1 was modified to specify that analysis of ORR using BIRC or investigator assessments will treat all subjects assessed as having non-measurable disease at baseline as non-responders.
- The definition of DCR in Section 3.4.3.3 was modified to state that single responses of CR and non-CR/non-PD in subjects with non-measurable disease at baseline will be eligible to contribute to disease control.
- The definition of the Measurable iCNS disease population (Section 3.2.5.3) was changed from being a subset of the ITT population to being a subset of the all iCNS disease population
- Clarification below Table 3 that subjects assessed as having non-measurable disease at baseline will be summarized as non-responders, and that patients with no valid time point responses will also be reported as non-responders

- Specification in Section 3.4.3.2 that censoring for anticancer therapies recorded during safety follow up will occur as of the date of contact at which the first occurrence of therapy is noted
- The definition of time to response analyses in Section 3.5.3.2 was simplified to the presentation of summary statistics restricted to responders. Time to event models and figures were eliminated.
- Time to onset and duration were removed from the analyses of AEs in special categories in Section 3.7.1.1, and specification that incidence for each constituent PT was added.

Version 4 update was made to this plan on August 19, 2019, which included the following:

- The alpha levels allocated to IA1, IA2, and FA were corrected based on O'Brien Fleming Lan-DeMets approach (this was erroneously reported in the protocol and previous version of the SAP as it was inadvertently calculated using the O'Brien-Fleming boundary.) in Section 3.1 and 3.4.3.6.

- Based on Intent-to-Treat Principle, subjects who were randomized but did not take any study treatment should be analyzed the same way as any other randomized subjects. Therefor, censoring rule for PFS and DOR were modified as below:

  The following sentence in Section 3.4.3.2 was removed.

  "Randomized subjects who did not receive any of their assigned study drug will be censored at follow-up Day1."

  The corresponding censoring rule #1 in Table 2 was removed. As a result of this change, subjects who were randomized but untreated will be analyzed based on the remaining progression and censoring rules in Table 2. The same change applies to Table A3-1 and Table A3-2.

  This change applies to the analysis of Duration of Response (DOR), intracranial Progression-free Survival (iPFS), and intracranial Duration of Response (iDOR) as well.

- Based on RECIST criteria v1.1, subjects who only had non-measurable disease at baseline should be counted as a responder if they had an overall response of CR. The following sentence in Section 3.5.1:

  "Any subjects who are identified as having only non-measurable disease at baseline by one of these assessment mechanisms (BIRC or investigator assessment) will not be eligible to be counted as responders (i.e. overall response of PR or CR) in analysis of results assessed by that mechanism, unless explicitly stated is exploratory analysis"

  was changed to

"Any subjects who are identified as having only non-measurable disease at baseline by one of these assessment mechanisms (BIRC or investigator assessment) will not be eligible to be counted as responders (i.e. overall response of PR or CR) in analyses of results assessed by that mechanism, unless their overall response is CR."

Same update made to the statement below Table 3 that subjects assessed as having non-measurable disease at baseline will be summarized as responders only if their overall response is CR, otherwise will be reported as non-responders.

According to protocol amendment 3, updates (version 5) were made to this plan on May 15, 2020, which included the following:

- The primary endpoint, PFS assessed by BIRC, was met at the first pre-planned interim analysis and confirmed at the second pre-planned interim analysis, before 198 events (PFS) events were observed.

- The timing of the primary analysis of BIRC assessed PFS in section 3.4.3.6 was modified.

  "The final analysis of the primary endpoint will be performed after 198 events have been observed, and will be tested at a 2-sided alpha level of 0.044."

  was changed to

  "The primary analysis of the primary endpoint will be performed after 198 events have been observed or at End-of-Study, whichever comes first, and will be tested at a 2-sided alpha level of 0.044."

- The timing of OS analysis in section 3.4.3.6 and 3.5.2.1 were modified.

  "Analyses of OS will also be performed at the time of the interim and final analysis, if the primary endpoint is met."

  was changed to

  "Analyses of OS will also be performed at the time of the interim analysis and at End-of-Study, if the primary endpoint is met."

"The primary assessment of the first three endpoints above will be performed at the time when the assessment of the primary endpoint is completed, either when the efficacy boundary is surpassed on one of the interim analyses or at the time the final analysis is done.

The primary assessment of OS will be performed when 198 PFS events to perform the final analysis of the primary endpoint have been observed, or at final database lock if the study is terminated prior to this occurs prior to the observation of 198 PFS events."

was changed to

"The primary assessment of all these four endpoints above will be performed at the time when the assessment of the primary endpoint is completed, either when the efficacy boundary is surpassed on one of the interim analyses or at the time the primary analysis is done."

2 STUDY DESIGN

This is a phase 3, randomized, open-label, comparative, multicenter, international study in which anaplastic lymphoma kinase positive (ALK+) non-small cell lung cancer (NSCLC) patients who have not previously received an ALK-targeted tyrosine kinase inhibitor (TKI) will be randomized in a 1:1 fashion to receive brigatinib (Arm A) or crizotinib (Arm B).

Patients will be stratified by the presence of central nervous system (CNS) metastases at baseline (Yes versus No) and prior chemotherapy used for locally advanced or metastatic disease (Yes versus No). For the purposes of stratification, prior chemotherapy is defined as completion of ≥1 full cycle of chemotherapy in the locally advanced or metastatic setting. An estimated 270 patients (135 in Arm A, 135 in Arm B) will be enrolled at approximately 150 centers.

Patients will be treated until they experience progressive disease (PD), intolerable toxicity, or another discontinuation criterion is met. Continuation of brigatinib beyond progression is permitted, at the investigator's discretion, if there is evidence of continued clinical benefit. Crossover from Arm B (crizotinib) to brigatinib is also permitted, at the investigator's discretion with the sponsor's medical monitor approval, for patients who have experienced objective progression determined by the blinded Independent Review Committee (BIRC).

Throughout the study, Adverse Events (AEs) will be assessed and categorized by the US National Cancer Institute Common Terminology Criteria for Adverse Events (see protocol Appendix A). Patients will be evaluated according to the Schedule of Events in protocol Section 11.1.

2.1 Randomization

Patients will be randomized in a 1:1 ratio to receive either brigatinib 90 mg once daily (QD) orally for 7 days followed by 180 mg QD orally continuously (Arm A) or crizotinib 250 mg twice daily (BID) orally (Arm B). Patients will be stratified by the following two factors, each having two levels:

1. intracranial Central Nervous System (iCNS) metastases at baseline (Yes versus No)
2. Prior chemotherapy use for locally advanced or metastatic disease (Yes versus No). For the purposes of stratification, prior chemotherapy is defined as completion of ?1 full cycle of chemotherapy in the locally advanced or metastatic setting.

Specific instructions for randomization will be supplied in the Study Reference Manual. Randomization procedures should be performed following completion of eligibility assessments and prior to the initiation of treatment, and required approval by the Sponsor's Medical Monitor. This study is unblinded; patients, investigators, and the sponsor will know the identity of each patient's study drug assignment.

2.2 Study Objectives

2.2.1 Primary Objective

The primary objective of the study is to compare the efficacy of brigatinib to that of crizotinib in ALK+ locally advanced or metastatic NSCLC patients naive to ALK inhibitors, as evidenced by PFS.

2.2.2 Secondary Objectives

The secondary objectives of the study are:

1. To compare the efficacy of brigatinib to that of crizotinib, as evidenced by confirmed objective response rate (ORR), time to/duration of response, disease control rate (DCR), and Overall Survival (OS)
2. To compare the efficacy in the CNS of brigatinib to that of crizotinib, as evidenced by intracranial response and intracranial PFS in those patients with iCNS metastases at baseline
3. To assess the safety and tolerability of brigatinib in comparison with crizotinib
4. To determine pharmacokinetic (PK) parameters of brigatinib through population PK modeling
5. To assess patient-reported symptoms and health-related quality of life (HRQoL) with the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ)-C30 (v3.0) in patients treated with brigatinib compared to those treated with crizotinib)

2.2.3 Exploratory Objectives

CCI

[redacted]

2.3 Study Endpoints

2.3.1 Primary Endpoint

The primary endpoint for this study is PFS, as assessed by the BIRC, per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 (Eisenhauer et al, 2009).

2.3.2 Secondary Endpoints

1. Confirmed ORR, as assessed by the BIRC, per RECIST v1.1
2. Confirmed intracranial ORR, as assessed by the BIRC
3. Intracranial PFS, as assessed by the BIRC
4. OS

The four endpoints above will be evaluated in a closed testing procedure as described in Section 3.5.2.1.

5. Duration of response, as assessed by the BIRC
6. Time to response, as assessed by the BIRC
7. Disease control rate, as assessed by the BIRC
8. Safety and tolerability
9. Change from baseline scores in global health status/quality of life (QOL) assessed with the EORTC QLQ-C30 (v3.0), and time-to-deterioration in dyspnea assessed with the EORTC QLQ-LC13 (v3.0)

2.3.3 Exploratory Endpoints

CCI [redacted]

3 GENERAL CONSIDERATIONS

The interim and final primary analyses will be conducted when the required number of PFS events have been observed as specified in protocol Section 15.5.2, and levels for statistical significance will be tested at the 2-sided alpha levels described in Section 3.4.3.5. Assessment of key secondary endpoints will be assessed in a closed testing procedure as described in Section 3.5.2.1. Comparisons between treatment groups for all other analyses will be tested at an unadjusted two-sided alpha level of 0.05, with two-sided 95% confidence intervals when appropriate unless specified otherwise.

Incidence of PFS events will be tracked throughout the course of the study, and the expected time of the required events for interim and final analyses will be estimated periodically. As the number of events approaches each of the pre-specified PFS event targets a Data Cut Off (DCO) date will be prospectively set. This date will be selected with the intent of reasonable assurance that a sufficient number of PFS events will be available.

All data collected will be included in the tabulation (SDTM) datasets, but data included in the analysis (ADaM) datasets will be truncated at the DCO date.

Unless otherwise stated baseline values in the main study period are defined as the last valid values collected during the time interval from the screening visit to the first dose date (time) of the randomized study treatment. Sites were instructed to perform C1D1 assessments prior to randomization and exposure to randomized study treatment. Therefore, all assessments performed on C1D1 are eligible for use baseline values. Descriptive statistics (such as means, medians, standard deviations, and ranges for continuous data, and counts/percentages for categorical data) will be used to summarize subject characteristics, study treatment administration, efficacy, safety, pharmacokinetic parameters, and genetic status of biomarkers. Data will also be displayed graphically where appropriate.

For the purpose of reporting efficacy and safety at specific time points, e.g., 6-month PFS, a month will comprise 30.4375 days. In case that at least one treated subject has been randomized to dosing regimen of 90 mg QD to 180 mg QD but have never been unable to successfully escalated to 180 mg (either during the first 7 days or later), these subjects will still be included on all the planned analyses for this regimen and may also be analyzed as a separate group.

Incidence for AEs and other important events will be reported at the subject level, with all subjects in the risk set used as the denominator and all subjects with at least one qualifying event constituting the numerator. In safety analysis, minimum or maximum values collected after first exposure to study drug may be used for summaries and change from baseline summaries.

Statistical analyses will be performed using SAS Statistical Software. Specifications for creating datasets for review, presentation, and analysis of results will be maintained by PRA Health Sciences under the direction of Takeda Pharmaceuticals.

3.1 Determination of Sample Size

For the purposes of this sample size calculation, the median PFS for crizotinib is estimated as 10 months (Solomon et al, 2014). Approximately 270 subjects will be randomized in a 1:1 fashion to receive brigatinib or crizotinib. A total of 198 events (progression or death among the randomized subjects) will provide 90% power to detect a clinically meaningful 6-month improvement in PFS (hazard ratio=0.625). This power projection is based on a 2-sided log-rank test, and is controlled at the 2-sided 0.0044 level, adjusting for the proposed interim analysis plan. The number of events is fixed, but the enrollment number may change based on an assessment of the overall event rate pooled across treatment groups (prior to the close of enrollment).

3.2 Analysis Populations

3.2.1 Intent–to-Treat (ITT) Population

The ITT population includes all subjects randomized to either regimen. The primary analyses of efficacy will be based on the ITT population.

3.2.2 Treated Population

The treated population for each regimen includes all subjects receiving at least one dose of study drug. Safety will be analyzed using the treated population.

3.2.3 FDA-Approved ALK Test Population

Section 15.5.2 of the protocol states a sensitivity analysis of the primary endpoint will be performed in the "FDA approved ALK test population". This population will consist of all subjects in the ITT population who also meet one of the two following criteria.

1) Were determined to be ALK+ by one of the tests identified in Inclusion Criterion 2a
2) Provided adequate tissue for central testing and were determined to be ALK+ at the central laboratory by the FDA-approved test

3.2.4 Per-protocol Population

The per-protocol population will be used for sensitivity analyses in order to characterize the effects of the two treatment regimens when used as intended according to the study protocol. Considerations for inclusion include consistency with the intended treatment population, adequate exposure to randomized treatment regimen, and sufficient post-exposure disease assessment to determine presence or absence of progressive disease. Per-protocol subjects will include all subjects in the FDA-Approved ALK Test Population who met the following criteria at study entry:

- Have histologically or cytologically confirmed stage IIIB (locally advanced or recurrent and not a candidate for definitive multimodality therapy) or stage IV NSCLC.
- Have at least one measurable lesion per RECIST v1.1 as assessed by the BIRC in the assessment of the baseline scan. Presence of measurable disease is defined as inclusion of at least one target lesion in the baseline RECIST assessment.
- Have at least one adequate post-baseline radiographic response assessment unless the reason is death or early discontinuation due to disease progression.

Usage of any systemic antineoplastic therapy that is prohibited by the protocol prior to BIRC-determined PD, death or discontinuation due to AE will disqualify a subject from the per- protocol population. The list of therapies that will be considered disqualifying for per-protocol population will be determined prior to the first interim analysis and updated prior to each evaluation of the primary endpoint. Use of radiotherapy consistent with the following condition from Section 13.2 of the protocol will not exclude a subject from the per-protocol population:

> Patients with CNS lesions requiring local radiotherapy such as stereotactic radiosurgery (SRS) are allowed to continue study drug after appropriate interruption, as determined by the investigator with sponsor agreement; however, for analysis purposes, these patients will be considered to have PD.

This definition of the per-protocol population is intended for sensitivity analysis of the primary endpoint and is only directly applicable for analysis of endpoints based on the main BIRC. If per-

protocol analyses of endpoints based on assessments performed by the investigator or iCNS BIRC are performed, different versions of the definition will be required.

3.2.5 Special Populations Specific to Baseline iCNS Disease status

See Section 3.4.1.2 of this document for a description of the iCNS BIRC process.

3.2.5.1 All iCNS Disease Population

The all iCNS disease population will consist of those subjects in the ITT population who were determined by the iCNS BIRC to have iCNS metastases at baseline regardless of whether they had at least one lesion that qualified as a target lesion in their baseline assessment.

3.2.5.2 No iCNS Disease Population

The no iCNS disease population will consist of those subjects in the ITT population who were not determined by the iCNS BIRC to have iCNS metastases at baseline.

3.2.5.3 Measurable iCNS Disease Population

Measurability of lesions is a core component definition of a potential target lesion in the RECIST v1.1 process and is retained in the modified RECIST used for iCNS disease assessment (Section 3.4.1.2). Therefore the measurable iCNS disease population will consist of those subjects in the all iCNS Disease population who were determined by the iCNS BIRC to have had at least one target lesion in their baseline assessment.

3.2.5.4 Non-Measurable iCNS Disease Population

The non-measurable iCNS disease population is intended to characterize subjects who were determined to have iCNS disease at baseline but did not have measurable lesions. This means that subjects with both measurable and non-measurable lesions at baseline will not be included in this population. Therefore the non-measurable iCNS disease population will consist of all subjects in the all iCNS disease population who are not included in the measurable iCNS disease population.

3.2.5.5 Active iCNS Disease Population

An active brain lesion is defined for the purpose of this study as meeting either of the following criteria:

1) A lesion that has not previously been irradiated
2) Having had prior radiation treatment but then having definitely progressed, as assessed by the investigator, after being irradiated.

However, the independence between disease assessments performed by the BIRC and those performed by the investigators requires limits on the amount and types of information transferred

between sites and the central reading process. Information as to whether any radiotherapy to the brain prior to study entry was made available to the BIRC readers, but this did not include whether the therapy was targeted at individual lesions or whether any irradiated lesions subsequently progressed. Therefore, assessment as to whether iCNS lesions are active cannot be done using BIRC time point responses. It can, however, be determined using the RECIST v1.1 assessment performed at baseline by the investigator since those assessments include lesion-level data on radiotherapy and subsequent progression.

Therefore the active iCNS disease population is defined as including all subjects in the all iCNS disease population who have one active iCNS lesion in their baseline RECIST v1.1 scan performed by the investigator. Analysis of efficacy in the active iCNS disease population will still be performed using data from the modified RECIST assessments performed by the BIRC.

3.2.5.6 Active Measurable iCNS Disease Population

The active measurable iCNS disease population is defined as all subjects included in both the measurable iCNS disease population and the active iCNS disease population.

3.2.5.7 Active Non-Measurable iCNS Disease Population

The active measurable iCNS disease population is defined as all subjects included in both the non-measurable iCNS disease population and the active iCNS disease population.

3.2.6 Crossover Population

Subjects in Arm B (crizotinib) who cross over to brigatinib following BIRC-assessed PD will have their data following start of brigatinib treatment analyzed as a single treatment group. Crossover must occur as directed in the protocol and result in the receipt of at least one dose of brigatinib. Subjects receiving brigatinib through alternative sources, such as by prescription in areas where it is commercially available, will not be included in the crossover population.

Indirect comparisons to other treatment experiences, such as the start of brigatinib in the Arm A subjects or the start of crizotinib within the same population, may be made but will usually not include formal statistical assessment of group differences.

Baseline values in the crossover period are defined as the last valid value prior to exposure to brigatinib for crossover subjects. Since no randomization date will be available for the crossover period the reference date for calculation of Study Day in the crossover period will be the date of first exposure to brigatinib. If a subject receives therapy qualifying as a censoring event in the period following the scan used as the baseline in the crossover period but prior to first exposure to brigatinib then all scans in the crossover period will be censored in the main analyses of crossover endpoints. Rules for determination of treatment emergence for AEs are described in Section 3.7.1.3.

Statistical comparison of group differences in OS that account for the effects of changing treatments will be performed as described in Section 3.5.2.5.

3.3 Demographics, Baseline Characteristics, and Subject Disposition

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

Subjects who cross over from crizotinib to brigatinib will be identified as having discontinued randomized therapy due to PD in the main disposition table, and as entered on the end of treatment page in the crossover disposition table.

Table 1 Demographics and Baseline Characteristics

Parameters	Categories for categorical parameters
From **Demographics** eCRF	
Age (continuous)	
Age (categorized in reference to age in whole years)	18-64, ≥65
	18-49, 50-64, 65-75, and ≥75
Gender	Male, Female
If Female, Is the patient of childbearing potential?	Yes, No
Race and Ethnicity	Individual responses will be presented as recorded and grouped based on observed responses
Geographic region/investigative sites	North America, Europe, Asia-Pacific
From **Eligibility Criteria** eCRF	
iCNS metastases at baseline, as randomized	Yes, No
Prior chemotherapy, as randomized	Yes, No
Strata, as randomized	iCNS metastases at baseline/Prior chemotherapy iCNS metastases at baseline/No prior chemotherapy No iCNS metastases at baseline/Prior chemotherapy No iCNS metastases at baseline/No prior chemotherapy
iCNS metastases at baseline, current	Yes, No
Prior chemotherapy, current	Yes, No
Current strata	iCNS metastases at baseline/Prior chemotherapy iCNS metastases at baseline/No prior chemotherapy No iCNS metastases at baseline/Prior chemotherapy No iCNS metastases at baseline/No prior chemotherapy
From **Diagnosis** eCRF	
Stage at initial diagnosis	IA, IB, IIA, IIB, IIIA, IIIB, IV, Unknown or not staged
Stage at study entry	IIIB, IV, Other
Time since initial diagnosis (continuous)	
Time since locally advanced or metastatic stage diagnosis (continuous)	
Histopathological classification at study entry	Adenocarcinoma, Adenosquamous Carcinoma, Large cell, Squamous, Unknown, Other
Lung involvement at Study Entry	Left Lung, Right Lung, Both Lungs, Lungs not involved
Other Organ Involvement at Study Entry	Yes, No Individual responses will be presented as recorded and grouped based on observed responses

From **Substance Use Cigarettes** eCRF	
Cigarette smoking history	Never, Former, Current, Unknown
Current cigarette smoking amount	Less than 20 per Day, 20 To 60 per Day, More Than 60 per Day
Former cigarette smoking amount	Less than 20 per Day, 20 To 60 per Day, More Than 60 per Day
From **Vital Signs** eCRF[1]	
Weight (continuous)	
Height (continuous)	
BMI (continuous)	Calculated from height and weight
From **ECOG Performance Status** eCRF	
Eastern Cooperative Oncology Group (ECOG) Performance Status	0, 1, 2
From **ALK Molecular Characterization** eCRF	
ALK status assessed	Yes, No
Parameters	**Categories for categorical parameters**
ALK assessed by FDA-approved test	Yes, No Defined as inclusion in the 3.2.3 FDA-Approved ALK Test Population (Section 3.2.3).
ALK rearrangement detected	Yes, No
From **Prior Radiation Therapy** eCRF	
Any prior radiotherapy	Yes, No
Best response to any radiotherapy	Complete Response Partial Response Stable Disease Progressive Disease Unable to Assess Unknown Other **Note:** select first above when > 1 radiotherapy listed
Prior radiotherapy to the brain	Yes, No
Best response to radiotherapy to the brain	Complete Response Partial Response Stable Disease Progressive Disease Unable to Assess Unknown Other **Note:** select first above when > 1 radiotherapy listed
From **Prior Anticancer Surgery** eCRF	
Prior Surgery	Yes, No
Diagnostic surgery	Yes, No
Therapeutic surgery	Yes, No
Purpose of surgery unknown	Yes, No
Purpose of surgery other	Yes, No
Resection Type	Complete Resection Partial Resection Unresectable Unknown Other Not Applicable **Note:** select first above when > 1 surgery listed

From **Prior Anticancer Therapy** eCRF	
Any prior anticancer therapy	Yes, No
Was at least 1 full cycle of any prior anticancer therapy administered in the locally advanced or metastatic setting	Yes, No
Any prior chemotherapy	Yes, No Agents will be identified as chemotherapy agents by a combination of programming logic and medical review
Was at least >1 full cycle of any prior anticancer therapy administered in the locally advanced or metastatic setting	Yes, No If at least one agent in a regimen is identified as chemotherapy this will be the response to the question about a full course for that regimen
Most recent systemic therapy	Chemotherapy, Other
Number of prior systemic anticancer therapies/regimens	0, 1, 2, etc.

[1] Values selected for each subject will be consistent with definition of baseline observation in the introduction to Section 3.

The list of characteristics described in the table above will also be used as the basis for sensitivity subgroup analyses of efficacy endpoints.

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

When analyses include the use of the factors used in stratification the current values will be used unless otherwise specified.

Analyses of most recent therapy will define most recent by the end date of therapy. If more than one agent is used in a regimen the end date for that regimen will be the last exposure date of any agent in the regimen.

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

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

Diagnosis Dates

- If day is missing but month and year are non-missing (YYYY-MM-UU), impute as the earliest of:
    - YYYY-MM-01
    - Randomization date

- If day and month are missing (YYYY-UU-UU), impute as the earliest of:
    - YYYY-01-01
    - Randomization date
- If month is missing (YYYY-UU-DD), but day and year are non-missing, ignore the day and impute as the earliest of:
    - YYYY-01-01
    - Randomization date

Prior Anti-Cancer Therapies Dates Start Date

- If day is missing but month and year are non-missing (YYYY-MM-UU), impute as the earliest of:
    - YYYY-MM-01
    - Randomization date
- If day and month are missing (YYYY-UU-UU), impute as the earliest of:
    - YYYY-01-01
    - Randomization date
- If month is missing (YYYY-UU-DD), but day and year are non-missing, ignore the day and impute as the earliest of:
    - YYYY-01-01
    - Randomization date

If after applying above rules, any prior treatment start dates are **before** diagnosis date, impute as diagnosis date.

Stop Date

- If day is missing but month and year are non-missing (YYYY-MM-UU), impute as the earliest of:
    - The last day of the month: YYYY-01-31, YYYY-02-28 or YYYY-02-29, YYYY-03-31, YYYY-04-30, YYYY-05-31, YYYY-06-30, YYYY-07-31, YYYY-08-31, YYYY-09-30, YYYY-10-31, YYYY-11-30, YYYY-12-31
    - Randomization date
- If day and month are missing (YYYY-UU-UU), impute as the earliest of:
    - YYYY-12-31
    - Randomization date

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

Duration of a selected prior anti-cancer therapy will be calculated using the following formula: duration = stop date – start date. Duration will be calculated with imputed dates when necessary. If a subject is randomized after more than one regimen of chemotherapy, contrary to the entry criterion of no more than one regimen, only the duration of the most recent regimen will be reported and summarized.

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

3.4 Efficacy Analysis

3.4.1 Disease Assessment

3.4.1.1 RECIST Assessment

Radiological scans for the assessment of disease progression and tumor response are scheduled to be performed every 8 weeks, and unscheduled scans can be performed at other times. Images from all scans performed from screening through the end of study participation will be transferred to the BIRC and evaluated per RECIST v1.1 (Eisenhauer, 2009) as described by the BIRC charter and associated documentation. The BIRC process is designed to account for commonly encountered situations in the evaluation of NSCLC severity in human trials, such as scans that cannot be evaluated and radiological assessments for a single visit performed on different days (e.g., computed tomography (CT) and magnetic resonance imaging (MRI)).

Validity of the findings of the BIRC will be enhanced by using two primary readers supported by an adjudicator; having readers of a scan blinded to all subsequent scans; keeping each reader blinded to the findings of the other reader; maintaining data through a transparent and documented data management process; and, most importantly, keeping all members of the BIRC blinded to study treatment and any other information collected on subjects following randomization other than the scans themselves.

Evaluable scans are defined as scans that, after completion of the BIRC assessment process, have a value for response that indicates a RECIST assessment of complete response (CR), partial response (PR), stable disease (SD), or PD. In cases where a subject was assessed as having non- measurable disease at baseline a time point response of non-CR/non-PD will be accepted as evaluable. Handling of scans of subjects with non-measurable disease at baseline in response analyses is described in Section 3.5.2.1. Unevaluable scans will be assigned not evaluable (NE) and not be used for censoring.

Expedited review of scans will continue until first occurrence of BIRC-determined disease progression.

For subjects who discontinue the study treatment due to a reason other than BIRC-determined progressive disease, additional tumor assessments should be documented, if available, until disease progression per BIRC or start of another systemic anti-cancer therapy. For subjects in Arm A who

continue brigatinib beyond disease progression, tumor assessments will continue to be performed every 8 weeks. For subjects in Arm B who crossover to brigatinib after BIRC determined progression on crizotinib, the scan at which BIRC-determined progression was determined will also be re-evaluated (per RECIST v1.1 criteria) as a new baseline for the Crossover Phase unless >21 days prior to the start of brigatinib therapy (Crossover C1D1), which will require a new disease assessment prior to starting brigatinib.

RECIST assessment will also be performed by the investigator.

3.4.1.2 Modified RECIST Assessment of iCNS Disease Burden

Intracranial CNS (iCNS) disease burden will be assessed using a modified RECIST. These assessments will be performed by a different BIRC team of primary readers and adjudicator(s) that will be conducted as consistently with the process for the primary endpoint as possible. The important modifications in for this assessment are:

1. Standard RECIST requires that subjects have measurable disease defined as the presence of at least one suitable target lesion. The modified iCNS assessments will specify how to handle subjects whose baseline iCNS disease does not meet the criteria for measurability. The assessment of baseline status will consist of one of three determinations:
    a. No iCNS disease – these subjects will be followed up for new lesions in the brain
    b. iCNS disease present but not measurable – in this case the readers will be instructed to enter as many non-target lesions as possible.
    c. Measurable iCNS disease – subjects in this category must have at least one target lesion (i.e., ≥ 10 mm in the longest diameter).
2. Standard RECIST allows up to 5 target lesions but specifies that no more than two target lesions can be in the same organ system. The modified iCNS assessment allows up to 5 target lesions in the brain
3. The modified RECIST will generate the following levels of response
    a. All non-readable scans will be assigned response of NE
    b. Subjects without baseline brain metastases disease per iCNS BIRC can have time point responses of PD or No Disease (ND)
    c. Time point responses for subjects with measurable disease at baseline will consist of CR, PR, SD, and PD
    d. Time point responses for subjects with only non-measurable disease at baseline will consist of CR, non-CR/non-PD, and PD

For the analysis of intracranial CNS ORR and PFS: If a subject progresses due to lesions outside the intracranial CNS and continues on study treatment, this subject will continue to be evaluated as SD, PR, or CR in the intracranial CNS until progression in the iCNS or discontinuation from the study treatment.

3.4.2 Definition and Analysis of the Primary Efficacy Endpoint

3.4.2.1 Definition of the Primary Efficacy Endpoint

The primary endpoint, PFS assessed by BIRC, is defined as the time interval from the date of randomization until the date of the PFS event as defined below. It will be censored for subjects who have not had a PFS event. Follow-up for PFS will also be censored in the primary efficacy analyses upon documentation of certain other events (such as initiation of disallowed therapy) as described in Table 2 below.

3.4.3 Primary Efficacy Endpoints Analysis

Important aspects of the definition of the PFS event and censoring are described in this section. The detailed scheme of progression and censoring for the primary analysis of PFS is specified in Table 2. These rules are defined consistently with FDA Guidance: "Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics Guidance for Industry" (April 2015).

3.4.3.1 Data handling Rules for the Primary Analysis of PFS

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

- A PFS event is defined as death or BIRC-determined RECIST progression. Follow-up for any subject with a PFS event will be uncensored in the analysis of the primary endpoint unless one of the censoring rules in Section 3.4.3.2 takes precedence.
- The PFS date is the earliest date of death or first observation of BIRC-determined RECIST progression events. All follow-up after the first PFS event will be excluded from the analysis of the primary endpoint
- The determination of the PFS event will include any PD observed by BIRC on or after randomization even if it does not occur at a scheduled study visit or is observed after the End of Treatment visit
- The date of progression is defined as the date at which disease progression was first evident and is determined by the BIRC readers in the BIRC system prior to transfer of data to the sponsor.
- Subjects may continue the study treatment or switch from crizotinib to brigatinib beyond documented progressive disease per RECIST v1.1 at the investigator's discretion. Among these subjects, only the date of the first documented evidence of progressive disease will be used to define the primary endpoint of PFS.
- Scans performed after the End of Treatment visit that are assessed by BIRC will be included in the analysis of the primary endpoint as long as one of the censoring rules does not take precedence.
- Per Section 13.2 of the protocol, subjects receiving radiotherapy for iCNS metastases will be considered as having PD in the primary analysis at the time such therapy was initiated.

3.4.3.2 Censoring Rules for the Primary Analysis of PFS

It is central to the definition of any time-to-event analysis that, for any subject who did not experience the event of interest, follow-up time is treated as censored in the analysis. In addition, any situation that impairs the continuity of unbiased follow-up should usually be considered as a reason to censor follow-up for any subject affected. Therefore, the additional censoring rules described in this section will be applied in the primary analysis.

- Subjects who cannot be evaluated for progression due to missing or incomplete baseline scans or because they have no evaluable on-treatment scans will also be censored at follow-up Day 1.
- If a subject starts any other anti-cancer therapy prior to a PFS event then follow-up will be censored at the date of the last valid assessment prior to initiation of the anticancer therapy.
    - This condition does not apply to crizotinib subjects crossing over to brigatinib as defined in the study protocol since this event should only occur after BIRC-determined PD. However, if a subject does receive brigatinib prior to such a determination, follow-up for the primary endpoint will be censored at the time of first exposure to brigatinib.
    - The protocol does not allow for brigatinib subjects to switch to crizotinib while on study, so any such event will be treated the same as the start of any other anti-cancer therapy.
    - Anti-cancer therapies recorded during safety followup will be censored as of the date of contact at which the first occurrence of therapy is noted.
- If a subject misses two consecutive disease assessments then follow-up for the primary endpoint will be censored at the time of the last non-progressive scan prior to the missed interval. Disease assessments are scheduled 56 days apart, so allowing a two-week window around two missed visits yields a period of 126 days.
    - Therefore any progression or death occurring more than 126 days after the last non-progressive disease assessment will not be included as a PFS event in the primary analysis, and follow-up will be censored on the date of the last valid assessment.
    - Death occurring within 126 days of the last valid non-progressive scan will be included as a PFS event on the date of death
    - Though this rule is implemented to ensure that PFS follow-up included in the analysis of the primary endpoint is conducted in a manner consistent with the design of the study as reflected in the Schedule of Events (Table 3 of the protocol) adherence to the 126-day rule is not restricted to disease assessments performed at Day 1 of the odd cycles. Any unscheduled scan assessed by BIRC can contribute to PFS follow-up and restart the 126-day interval.
- Censoring events must occur at least one day prior to the PFS event date as defined in Section 3.4.3.1. For example, a subject could commence a therapy consistent with the censoring rules on the same day as BIRC-confirmed PD. In such a case the latter event

will take precedence and the subject's follow-up will not be censored in the primary analysis.

3.4.3.3 PFS Events and Censoring Using On-Study Anticancer Treatments

As noted in the previous two sections, receipt of anticancer therapy other than randomized treatment can be treated as a PFS event (Section 3.4.3.1) or cause follow-up time in the analysis to be censored (Section 3.4.3.2).

The process for selecting on-study therapies that will be treated as PFS or censoring events will start with a medical review of all therapies received by subjects on the study. A list of all the therapies will be compiled following DCO at each primary analysis. All occurrences of these therapies will be selected from the study data. An additional screen of concomitant medications will be used to identify antineoplastic agents. All therapies other than radiotherapy will be treated as potential censoring events, and the earliest such therapy for each subject will be utilized in the censoring algorithm. An example list of therapies to be treated strictly as potential censoring events is included in Appendix 3.

Radiotherapy entries will be checked for whether they were targeted at the brain and/or central nervous system. If so, the event will be treated a potential PFS event and treated as being the same as BIRC PD in the determination of time to PFS. Otherwise the event will be treated as a potential censoring event in the same manner as all other on-study anticancer therapies. An example list of therapies to be treated as either potential PFS or censoring events is included in Appendix 4.

The final versions of the lists used for PFS events or censoring will be included as appendices in the CSR.

3.4.3.4 Calculation of Follow-up Time for the Primary Analysis of PFS

For subjects who were observed to have a PFS event follow-up time is defined as:

Follow-up = PFS date – randomization date +1

which will be entered into the model as uncensored time.

For other subjects follow-up time is defined as:

Follow-up = Censoring date – randomization date +1

which will be entered into the model as censored time.

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

#Rule	Situation	Date of progression or censoring	Outcome
1	Missing or incomplete baseline	Day 1 (randomization date)	Censored
2	No valid post-exposure assessment	Day 1 (randomization date)	Censored
3	No measurable disease at baseline	Date of new lesion(s) or substantial worsening in	Progressed
		Date of last evaluable progression-free	Censored
4	a) Death, PD, or non-PD observed > 126 days after	Date of last evaluable progression-free radiographic assessment	Censored
	b) Death or PD ≤ 126 days after last valid RECIST	Earliest of date of death and date of 1st PD	Progressed
5	No progression or death	Date of last evaluable progression-free	Censored
6	Death before first PD assessment or between	Date of death	Progressed
7	Death after one missed	Handled according to rule #4	
8	Death after two or more missed radiographic	Handled according to rule #4	
9	New anticancer treatment started prior to PFS event	Date of last evaluable progression-free radiographic assessment prior to initiation	Censored
10.	Cancer-related surgery prior to documented disease	Date of last evaluable progression-free radiographic assessment	Censored
11	Disease progression documented between scheduled visits	Date of progression as reported by BIRC	Progressed
12.	Disease progression preceded by 1 missed follow-up	Handled according to rule #4	
13.	Disease progression preceded by 2 or more consecutive	Handled according to rule #4	

Note: Evaluable radiographic assessments used for censoring dates are restricted to BIRC assessments and exclude those by iCNS BIRC and investigator.

3.4.3.5 Evaluation of the Primary Efficacy Endpoint

The primary analysis of the primary endpoint will be performed using a 2-sided stratified log- rank test (stratification factors: presence of iCNS metastases at baseline [Yes versus No], and prior chemotherapy for locally advanced or metastatic disease [Yes versus No]) to compare the BIRC-assessed PFS of subjects randomized to brigatinib with the BIRC-assessed PFS of subjects randomized to crizotinib. The overall (2-sided) Type I error rate will be controlled at 0.05. The primary analysis will be based on the ITT population. Median PFS and 95% confidence intervals will be estimated for each treatment arm using the Kaplan-Meier method (Kaplan and Meier, 1958). Additionally, hazard ratios will be estimated using the Cox regression model with the stratification factors as covariates (Cox, 1972). The analysis will use the current values of the stratification factors.

3.4.3.6 Interim and Final Analyses of the Primary Endpoint

Two interim analyses are planned after approximately 50% and 75% of the total expected events (progression or death) have been observed. An O'Brien-Fleming Lan-DeMets alpha spending function (DeMets and Lan, 1994) will be used to control the overall alpha level at 0.05 2-sided.

The first interim analysis will be performed after the first 99 events have been observed. The primary endpoint of PFS will be tested at a 2-sided alpha level of 0.0031. A second interim analysis will be performed after 149 events and the primary endpoint will be tested at a 2-sided alpha level of 0.0183. The primary analysis of the primary endpoint will be performed after 198 events have been observed or at End-of-Study, whichever comes first, and will be tested at a 2-sided alpha level of 0.044. Analyses of OS will also be performed at the time of the interim analysis and at End-of-Study, if the primary endpoint is met. If the efficacy boundary is surpassed at an interim analysis, subjects will remain on assigned treatments unless a survival advantage is noted by the Data Monitoring Committee.

The first occurrence of a successful evaluation of the primary endpoint, defined as surpassing the critical value at a pre-planned analysis, will complete the inferential statistical evaluation of the primary endpoint for this study. All subsequent analyses after the conclusion of the primary analysis will be non-inferential, even for analyses conducted as specified in the protocol for subsequent analyses of the primary endpoint. Interpretation of additional analyses of PFS assessed by BIRC, as well as secondary and exploratory endpoints, must be informed by the context that they are being performed for a study for which the primary endpoint has already been met.

It is possible that, when database extracts are created for a pre-specified evaluation of the primary endpoint, there could be more PFS events than were specified in the protocol. If this happens the following procedure will be used to determine the PFS events included in the analysis:

1) The eligible PFS events will be sorted in the order they occurred according to calendar date. PFS events will not be eligible in this context if they are preceded by censoring events such as disallowed anticancer therapy.
2) As stated in Section 3, a Data Cut Off (DCO) date will be prospectively set at the estimated time of the pre-specified event counts for the interim and final analyses.
3) All PFS events that occurred on or before that date will be included in the analysis

4) Follow-up for all other subjects will be censored for PFS analyses at the time of the last BIRC evaluation prior to that date unless other censoring rules take precedence as described in step 1.
5) The number of PFS events in an evaluation of the primary endpoint could be different from the number specified in the protocol. If this occurs it will be noted in the CSR.

If either the Cox model or log-rank test report convergence failures or unstable parameter estimates the primary analysis will be repeat by dropping one of the stratification factors:

- First the model will drop iCNS metastases at baseline.
- If this does not produce a stable model then drop prior chemotherapy (while retaining baseline iCNS metastases from the initial model). If issues persist then drop both of the stratification parameters.
- If the model still does not converge after dropping both stratification factors the Cox model estimates will be reported as NE.

3.4.3.7 Protocol-Specified Sensitivity Analyses of the Primary Endpoint

Protocol Section 15.5.2 states that the following sensitivity analyses will be performed. Sensitivity analyses of the primary endpoint of PFS will also be performed in the following populations:

- Per-protocol population, as assessed by the BIRC
- FDA approved ALK test population, as assessed by the BIRC
- ITT population (all subjects randomized to a treatment arm), as assessed by the investigator.

Subgroup analyses will be performed by baseline potential prognostic factors.

3.4.3.8 Additional Sensitivity Analyses of the Primary Endpoint

Sensitivity analyses will be performed to test the effects of specific aspects of the study design, conduct, and data handling on the robustness of the conclusions reached in the analysis of the primary endpoint. Two major sensitivity analyses are detailed in tables shown in Appendix 2.

- Sensitivity Analysis 1 (Table A3-1) – Censor at date of last treatment administration if PFS event not observed prior to the End of Treatment Visit. Other censoring rules from the primary analysis of PFS remain in effect.
- Sensitivity Analysis 2 (Table A3-2) – Do not censor PFS events for reasons of time lapse from last assessment (i.e., no 126-day rule) or usage of disallowed therapies.

Each sensitivity analysis is targeted at specific changes to the rules for the primary analysis, and any rules not specified in the sensitivity analysis will remain as specified in the primary analysis. For example, in Sensitivity Analysis 1 the date used for censoring will change but the censoring rules for disallowed therapies will remain in effect.

Sensitivity subgroup analyses will be performed on the primary endpoint using all baseline factors listed in Table 1 with the following considerations:

- The primary efficacy models will be repeated for each level of the subgroup variable restricted to the subjects at the level being examined. For example, the sensitivity subgroup analysis of gender will be run once restricted to female subjects and once restricted to males.
- The model run will be with the same terms and specifications as the primary model to which it refers.
    - For example, the sensitivity subgroup analysis of gender would include the treatment group and stratification variables used in the primary analysis.
    - Each sensitivity subgroup model must have adequate data to support a valid statistical analysis. Cases where data are determined to be inadequate will be noted in the CSR. Such cases will include but may not be limited to the following:
        - Any level of the subgroup variable with fewer than 10 subjects in either of the treatment groups
        - When appropriate, as for time-to-event models, any level in which none of the events in question were observed in either of the treatment groups
        - If the modeling procedure reports problems such as convergence failure or generation of unstable parameter estimates
    - In cases where the definition of the subgroup being tested precludes specification of the model as used in the primary analysis the change to the model will be noted. For example, when running sensitivity subgroup models broken out by prior chemotherapy the stratification by prior chemotherapy cannot be included in the model.
        - In cases where use of the primary model specification results in conditions suggesting inadequate data, but where elimination of one or both of the stratification variables addresses the data issue, then the simpler model will be used. Any modifications to the primary model will be noted.
- The main focus of sensitivity subgroup analyses will be the point estimates and confidence intervals of the main treatment effect between the levels of the subgroup variable.
    - In the primary analysis this would mean comparisons of median PFS time and hazard ratio between treatment groups within the levels of the subgroup variable.
    - Interpretation of the effect of the subgroup will not be based on a threshold of statistical significance, such as a p-value less than 0.05, but rather on interpretations of the estimates as compared to the primary models and to other level(s) of the subgroup variable. Particular attention will be paid to whether point estimates across levels of the subgroup variable are on the same side of the null value of the statistical model and whether there is substantial overlap between the corresponding confidence intervals.

- The effects of subgroups on the estimates of treatment effect will be examined graphically using Forest plots. Factors to be presented include current and as-randomized values of the stratification factors as well as Age (18-49, 50-64, 65-75, and ≥75), Race (Asian vs. non-Asian), Gender, Smoking Status (never, former, current, unknown), and ECOG Performance Status (0 vs. 1 or 2).

Summarization and reporting of the results of sensitivity analyses will be abbreviated as compared to that for the primary model unless important differences are noted.

3.5 Definitions and Analyses of Secondary Efficacy Endpoints

3.5.1 General Rules for Secondary Efficacy Endpoints

The rules outlined in this section apply to definition of response endpoints based on RECIST v1.1 evaluations performed by BIRC and the investigators as well as modified RECIST by the iCNS BIRC.

Entry criteria required measurable disease, as evidenced by at least one target lesion, at baseline. This criterion, according to study conduct and randomization procedures, could only be applied to investigator assessments, and it does not ensure that every subject would also be assessed as having measurable disease by the BIRC. It is also possible that after randomization all target lesions for a subject could have been reclassified as non-target lesions in the investigator's baseline assessment. Any subjects who are identified as having only non-measurable disease at baseline by one of these assessment mechanisms (BIRC or investigator assessment) will not be eligible to be counted as responders (i.e. overall response of PR or CR) in analyses of results assessed by that mechanism, unless their overall response is CR. This restriction will not be to analyses of iCNS BIRC assessments when subjects with non-measurable iCNS disease are included. Two types of sensitivity analyses of secondary endpoints will generally be performed:

- Sensitivity analyses using investigator assessments will be performed on BIRC-assessed RECIST endpoints.
- Analyses of all (confirmed + unconfirmed) responses will be performed as sensitivity analyses of confirmed responses will be performed unless the number of unconfirmed responses is low.

Other sensitivity analyses, such as by subgroups or through modifications of model specifications, may be performed but are not required by this plan.

3.5.1.1 Confirmation of Response

The primary analysis of any response endpoint for this study will be based on confirmed response, as assessed by BIRC, unless otherwise specified. Confirmation requires that an initial response be followed by observation of another response as good or better than the initial after a suitable interval of time, which would entail one of the following pattern:

- PR followed by PR or CR
- CR followed CR

Disease control rate (DCR) can be considered a special case of confirmed response. In this situation disease control is defined as at least one valid observation that the disease state on therapy has not progressed as compared to baseline. Therefore, DCR (or, confirmed SD) only requires one on-study assessment, whereas confirmed ORR requires two. In subjects with measurable disease at baseline DCR can be derived directly from best response meaning that any subject with a best response of CR, PR, or SD will be reported as having demonstrated DCR. For subjects without measurable disease at baseline subjects with best responses of CR (in non-target lesions) or non-CR/non-PD will also be reported as having demonstrated DCR. In iCNS analyses responses of CR (in non-target lesions) or non-CR/non-PD are eligible to be reported as responses according to the rules for response, best response, and ORR.

Censoring rules for PFS events will also apply to other RECIST-based endpoints. Time point responses following the first PD will be excluded, as will those observed after censoring events have occurred. Censoring rules based on observed RECIST assessments such as the exclusion of those observed after the first PD will restricted to those using the same assessment process. For example, BIRC assessments following the first BIRC-assessed PD will be excluded from consideration of confirmed response but will not be affected by iCNS BIRC or the investigator assessment. In addition, when a rule specifies use of the most recent valid scan date the set of dates to be used will be restricted to those of the same type (BIRC, iCNS BIRC, or investigator) as used in the analysis.

In general the determination of Best Confirmed Response for each subject will be based on the BIRC response patterns for that subject. However, response patterns cannot be interpreted without incorporating the timing of the assessments that constitute that pattern. The handling of response patterns will be described in this section, and the handling of timing will be specified in the next section.

A strict interpretation of the RECIST v1.1 algorithm would require handling CR followed by PR or SD as PD. If, following a time point response of CR, a reader enters any information indicating the presence of disease into the BIRC system they will be presented with a default time point response of PD. The reader will, however, have the ability to override this default if so guided by their clinical judgment. However, for situation 1 above to be observed in the BIRC data both the initial CR and the subsequent PR/SD would have to have agreement with the second reader or the adjudicator, which is expected to be extremely rare if observed at all. If this situation occurs in the observed data its impact on analyses will be assessed and noted in the CSR, with sensitivity analyses if warranted.

Another scenario that is reasonably likely to occur is PR followed by SD. This could occur if the sum of target lesions hovers around the 30% reduction threshold used in the definition of PR. Appendix III of the RECIST v1.1 guideline states a single SD can be observed in between the initial PR and its confirmatory assessment, but that more than one SD should cause confirmation of that initial response to fail. Such an event will not, however, prevent subsequent PRs from being confirmed.

A single scan that cannot be interpreted, resulting in the occurrence of a time point response of NE, that follows an initial response will not invalidate confirmation of that response. Therefore, the following patterns of response will be considered as confirmed CR and confirmed PR, respectively:

- CR-NE-CR
- PR-NE-[PR or CR]

More than one NE such as CR-NE-NE-CR, however, will preclude the initial response from being confirmed.

Additional considerations in the handling of time point responses of SD and NE in the confirmation of response are:

- In the confirmation of PR responses of SD and NE will be treated as equivalent. Therefore the pattern of PR-SD-NE-PR would cause the first PR to be unconfirmed. However, the second PR in the series above can be confirmed if subsequent responses are consistent with that interpretation.

- SD and NE are not equivalent in the confirmation of CR
- NE cannot contribute to the determination of DCR but will also not, on its own, invalidate observation of disease control on subsequent scans
- All assessments used for confirmation must meet the timing requirements described below

Any time point response of PD precludes consideration of any subsequent responses for confirmation of initial response. Early Death (ED) will only be reported when death occurs with no prior valid scan. The rules for handling confirmation of response are presented in Table 3:

Table 3 Response Confirmation Patterns

	Pattern	Confirmation of Initial Response?	Best Confirmed Response	Best Unconfirmed Response
1	CR-PD	N	SD	CR
2	PR-PD	N	SD	PR
3	SD-PD	N	SD	SD
4	NE-PD	N/A	PD	PD
5	PD	N/A	PD	PD
6	Early Death (ED)	N/A	ED	ED
7	CR-CR	Y	CR	CR
8	CR-NE-CR	Y	CR	CR
9	CR-SD-CR	N	SD	CR
10	CR-NE-NE-CR	N	SD	CR
11	PR-PR	Y	PR	PR
12	PR-CR	Y	PR	CR
13	PR-NE-PR	Y	PR	PR
14	PR-NE-CR	Y	PR	CR
15	PR-SD-PR	Y	PR	PR
16	PR-SD-CR	Y	PR	CR
17	PR-NE-NE-PR	N	SD	PR
18	PR-SD-SD-PR	N	SD	PR
19	PR-NE-SD-PR	N	SD	PR
20	PR-SD-NE-PR	N	SD	PR
21	PR-NE-NE-CR	N	SD	CR
22	PR-SD-SD-CR	N	SD	CR
23	PR-NE-SD-CR	N	SD	CR
24	PR-SD-NE-CR	N	SD	CR

Subjects with non-measurable disease at baseline will be reported as responders only if their overall response is CR. Otherwise, they will be reported as non-responders. Subjects with no evaluable scans will be reported as non-responders.

CCI

[redacted]

Per Section 13.2 of the protocol subjects receiving radiotherapy for iCNS metastases will be considered as having PD in the primary analysis at the time such therapy was initiated in the same manner as if a BIRC-assessed time point response of PD was observed at that point.

3.5.1.2 Timing of Disease Assessments Used for Confirmation of Response

The following statements from the protocol inform how the timing of assessments will be handled:

1) The definition of Disease Control Rate (Protocol Section 15.5.1) requires that "in the case of SD, criteria for SD must have been met at least once after randomization at a minimum interval of 6 weeks"
2) Imaging for possible confirmation of response "should be performed at least 4 weeks after initial response" (Protocol Section 11.1, Item #20)
3) The scheduling of disease assessments includes a window of ± 3days (Protocol Section 11.1, Item #20)

For DCR, six weeks would target the first eligible scan at Day 43 relative to randomization. Allowing a 3-day window around this target means that scans performed on or after Day 40 relative to randomization are eligible for consideration in the establishment of disease control for a subject. As described in the previous section the criteria for SD include unconfirmed PR and CR, but at least one of these responses must be observed on or after Day 40 to qualify for a Best Confirmed Response of SD and also for DCR. If such early responses of CR, PR, or SD are followed by PD then the subject's best confirmed response will be PD, otherwise it will be NE.

Early responses of CR and PR will be eligible to determine best unconfirmed response and will be considered to be eligible for confirmation. For the purposes of response confirmation, a confirmation interval will be defined as:

Scan interval = scan date 2 – scan date 1 + 1

Therefore, for the purposes of response confirmation the confirmatory scan would be targeted for Day 29 relative to observation of initial response. Allowing for the 3-day window means that any scan on or before Day 25 relative to observation of initial response cannot be used for confirmation of that response. However, to be consistent with the definition of DCR it will also be required that the confirmatory scan meet the same rule as for determination of best confirmed response of SD. Therefore, a confirmatory scan must meet all of the following criteria:

1) It must be part of a pattern leading to confirmed response in Table 3.
2) The scan date must be after day 25 relative to the initial response date.
3) The scan must also be on or after day 40 relative to randomization date.

In addition, using the logic specified in Section 3.4.3.2, any scan greater than 126 days after an initial response cannot be used for confirmation of that response. If a subject does not have at least one evaluable scan over any 126-day interval all scans after that interval will be excluded from the determination of response.

3.5.2 Key Secondary Efficacy Endpoints and Analyses

3.5.2.1 Rules Handling the Overall Type I Error Rate for Secondary Endpoints

Key secondary endpoints will be tested using a closed testing procedure to control the overall type I error rate at 0.05. Analysis of a key secondary endpoint will be considered significant if the test for that endpoint and comparisons of all other secondary endpoints with a smaller rank are significant at the two-sided 0.05 significance level.

Rank-ordering of key secondary endpoints:

1. Confirmed ORR, as assessed by the BIRC, per RECIST v1.1
2. Confirmed intracranial ORR, as assessed by the BIRC
3. Intracranial PFS, as assessed by the BIRC
4. OS

The primary assessment of all these four endpoints above will be performed at the time when the assessment of the primary endpoint is completed, either when the efficacy boundary is surpassed on one of the interim analyses or at the time the primary analysis is done.

3.5.2.2 Confirmed ORR, as assessed by the BIRC, per RECIST v1.1

Confirmed ORR, as assessed by the BIRC, will be performed in the ITT population. ORR will be calculated using the following formula:

ORR per BIRC = (#confirmed BIRC-based CR or PR) / (#Randomized subjects)*100%.

Sensitivity analyses of this endpoint using different analysis populations (Section 3.2) will be calculated in the same manner after first restricting subjects included in the denominator and eligible for inclusion in the numerator to those in the analysis population of interest.

Confirmed ORR will be analyzed with the Mantel-Haenszel test (using the stratification factors) on the ITT population to compare the proportion of subjects achieving objective response between the 2 treatment arms. The exact 2-sided 95% binomial confidence intervals will be calculated.

Sensitivity analyses to be performed on this endpoint will be:

1) ORR defined as all confirmed and unconfirmed responses as determined by BIRC
2) Confirmed ORR as determined by the investigator
3) ORR defined as all confirmed and unconfirmed responses as determined by the investigator

3.5.2.3 Confirmed intracranial ORR, as assessed by the BIRC

iCNS ORR will be calculated in the same manner as for RECIST v1.1 BIRC using confirmed time point responses from the iCNS BIRC. Confirmed iCNS ORR will be analyzed for the following subsets of subjects:

1) The measurable iCNS disease population
2) The non-measurable iCNS disease population
3) The all iCNS disease population

Each of these analyses will be accompanied by a sensitivity analysis restricted to the subset of subjects with active lesions at baseline. Responses of iCNS CR in subjects with non-measurable iCNS disease at baseline will be counted as time point responses in analyses of iCNS ORR as defined in Table 3.

Analysis will differ in that the iCNS ORR will be analyzed with the Mantel-Haenszel test using only the stratification factor of prior chemotherapy.

A sensitivity analysis of confirmed + unconfirmed iCNS ORR will be performed.

3.5.2.4 Intracranial PFS, as assessed by the BIRC

Intracranial PFS will be defined and analyzed in the same manner as for PFS used in the primary endpoint with the exceptions that the progression events used will come from the iCNS BIRC and the analysis will be restricted to subjects identified as having brain metastases at baseline in the randomization.

The primary evaluation of iCNS PFS will be performed in the all iCNS disease population, with a sensitivity analysis in the active iCNS disease population. Additional analyses will be performed in the no iCNS disease population, in which case PFS events would consist of either the appearance of new brain metastases or death. iCNS PFS will also be analyzed in the measurable iCNS disease population with a sensitivity analysis in the active measurable iCNS disease population.

3.5.2.4.1 Data handling Rules Specific to Secondary Analysis of iCNS PFS

Conduct and analysis of iCNS response and progression will occur independently of systemic assessment of disease. If a subject progresses due to lesions outside the iCNS and continues on their randomized study treatment, the subject will continue to be evaluated as SD, PR, or CR in the iCNS until disease progression in the iCNS or the subject discontinues randomized study treatment. Confirmation of CR or PR should be performed 4 weeks (allowing a minus 3-day window) or more after initial response.

The primary analysis of iCNS PFS will be based on radiographic assessments in the iCNS by the BIRC. The primary analysis will only include randomized subjects with active brain metastases at baseline. In all the analyses, the progression and censoring scheme in Table 2, Section 3.5.2.4.1 will be used except that the radiographic assessments are restricted to the disease in the CNS.

Per Section 13.2 of the protocol subjects receiving radiotherapy for iCNS metastases will be considered as having PD in the analysis of iCNS PFS at the time such therapy was initiated.

3.5.2.5 OS

OS is defined as the time interval from the date of randomization until death due to any cause in the ITT population. OS will be censored on the date of last contact for subjects who are still alive. Contact must provide positive evidence that the subject is alive at the time it is made. In the primary analysis of OS follow-up time will not be censored for initiation of non-study anticancer therapies, including crossover from crizotinib to brigatinib.

The primary analysis of OS will be conducted using the same statistical methods as for the primary endpoint.

Follow-up time in the primary analysis of OS will not be censored for any other reason. Sensitivity analyses using additional censoring rules such as initiation of disallowed anticancer therapies may be performed (see below).

For subjects who have been randomized and are not treated, the OS will be defined as the time interval from randomization date to death date or date of the last contact if available.

A majority of subjects can be expected to take subsequent anticancer therapy following disease progression, including crossover from crizotinib to brigatinib as directed in the protocol. To adjust for the effects of subsequent therapy, the following sensitivity analysis may be performed:

1) Primary analyses except additional censoring at the start day of subsequent anti-cancer therapy
2) Rank preserving structural failure time (RPSFT) analysis
3) Inverse probability of censoring weighted (IPCW) analysis

3.5.3 Additional Secondary Endpoints

3.5.3.1 Duration of response, as assessed by the BIRC

The analysis of response duration will be based on disease assessment by the BIRC and use the progression and censoring scheme in Table 2, Section 3.5.2.4.1. The analysis of response duration will only include ITT subjects with confirmed CR or PR. An additional analysis of duration of response will be performed using disease assessment performed by the investigator. Censoring in the analysis of duration of response will be the same as for PFS.

- Duration of confirmed response = PFS event/censoring date – date of first response + 1 in subjects whose response was confirmed
- Duration of response = PFS event/censoring date – date of first response + 1

Median values and 2-sided 95% confidence intervals will be estimated using Kaplan-Meier (KM) method (Kaplan and Meier, 1958) in the ITT population. The KM-estimated PFS rates and OS rates at

12 and 24 months and the associated 2-sided 95% confidence intervals will be computed. Duration of response will also be summarized with descriptive statistics for subjects with confirmed CR or PR and the Kaplan-Meier method in which follow-up for subjects without PFS events will be censored.

Figures associated with this endpoint will use the same rules as described above.

3.5.3.2 Time to response, as assessed by the BIRC

Time to response as assessed by the BIRC will be summarized for confirmed responders using descriptive statistics. The primary analysis of time to response will use the same definition of confirmed response as used for ORR in Section 3.5.2.2. The time of response will be the date of the initial response that was ultimately confirmed, and not the date at which the confirmatory response was observed.

- Time to confirmed response = Date of initial response – date of randomization + 1.

3.5.3.3 Disease control rate, as assessed by the BIRC

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

Responses of CR and non-CR/non-PD in subjects with non-measurable disease will be eligible for confirmation of DCR for BIRC, investigator, and iCNS BIRC assessments.

Disease control rate will be analyzed with the Mantel-Hanzel test (using the stratification factors) on the ITT population to compare the proportion of subjects achieving disease control.

3.5.3.4 Safety and tolerability

See Section 3.7 for descriptions of endpoints and analyses of safety information.

3.5.3.5 Subject-reported symptoms and HRQoL scores, assessed with the EORTC QLQ-C30 (v3.0) and QLQ-LC13

The main subject-reported outcomes (PRO) endpoints of interest will be the Global Health Status/QoL Scale, based upon Question 29 and Question 30 of the QLQ-C30 (v3.0), and the Dyspnea Scale, based on the QLQ-LC13 Questions 3-5. The QLQ-C30 will be scored according to the EORTC QLQ-C30 (V3) Scoring Manual (Fayers et al, 2001) including transformation of responses into scores for analysis and the handling of missing data. The EORTC QLQ-C30 (V3) Scoring Manual also includes instructions for scoring the QLQ-LC13. Analysis of the Global Health Status/QoL Scale over time may be performed using mixed effects models that include randomized treatment group and the stratification factors used in randomization.

Improvement in global health status, defined as an improvement from baseline of 8.33 points or greater in the Global Health Status/QoL Scale, will be analyzed using the Mantel-Haenszel test as performed for ORR. The primary evaluation of this response endpoint will be performed using the Cycle 3, Day 1 assessment. The analysis may be repeated at other time points, and a time to response model may also be performed.

Worsening on the Dyspnea Scale for a subject will be defined as a 50% decline from the baseline score. Time to worsening on the Dyspnea Scale will be analyzed using the same methods as for the primary analysis of PFS.

Analyses of QoL endpoints will include all subjects who have a baseline and at least one post-baseline assessment. Time-to-event analyses of QoL will censor follow-up at the last evaluable QoL assessment for subjects for whom the event of interest is not observed.

3.6 Exploratory Analyses

3.6.1 Definitions of Protocol-Specified Exploratory Efficacy Endpoints

Exploratory Endpoints are specified in Section 9.1.3 of the protocol.

CCI

[redacted]

CCI

CCI

CCI

[redacted]

3.7 Safety Analysis

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

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

Unless otherwise specified all safety analyses will be reported according to actual study treatment received.

3.7.1 Adverse Events

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

All AEs entered in the clinical database (including treatment-emergent and non-treatment- emergent AEs) will be listed in by-subject listings or available for review in appropriate datasets. The incidence rates of treatment-emergent AEs, as well as the frequency of occurrence of overall toxicity categorized by maximum toxicity grades (severity), will be described. In addition, treatment-emergent AEs will be summarized by causal relationship to study treatment (in the opinion of the investigator) and action taken on study treatment, including dose modifications, interruptions and

discontinuation. Serious treatment-emergent AEs (SAEs), both overall and by causal relationship to study treatment, will also be summarized.

Incidence rate for treatment-emergent AEs and AE groups in special categories, e.g., pulmonary adverse events overall and within first 7 days, will be summarized and also be adjusted with subjects' entire exposure to study treatment and exposure through the AE initial onset. AEs in special categories will also be summarized in terms of number of subjects with at least one event in the category and by number of subjects with at least one event for each constituent preferred term. The special categories to be analyzed will include at minimum those listed in Appendix 1, and will align with the risk profile of brigatinib and crizotinib. All deaths occurring during a period of exposure to randomized therapy or within a period of up to 30 days following discontinuation from randomized therapy will be summarized by AEs leading to death and causal relationship to study treatment. The 30-day rule will be superseded in cases of crossover from crizotinib to brigatinib when done in compliance with the protocol, in which case all deaths on or after the first dose of brigatinib and within a period of up to 30 days following discontinuation from brigatinib will be summarized in the crossover period. All deaths occurring later than 30 days but resulting from treatment related adverse event(s) will be summarized by AEs leading to death and causal relationship to study treatment.

CCI

[redacted]

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

If adequate information regarding the investigator's assessment of relationship to study treatment, the relationship to the study treatment will be imputed as possibly related to study drug for treatment-emergent AEs. This rule will be applied to events in both treatment arms.

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

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

Resolution Date

- If day is missing but month and year are non-missing (YYYY-MM-UU), impute as the earliest of:
    - Last day of the month (28, 29, 30 or 31 depending on in which month the adverse event resolved)
    - Data cutoff date
    - Death date

- If day and month are missing (YYYY-UU-UU), impute as the earliest of:
    - December 31 (YYYY-12-31)
    - Data cutoff date
    - Death date
- If date is completely missing (i.e. AE is ongoing), impute as earliest of:
    - Data cutoff date
    - Treatment discontinuation date + 30 days
    - Death date

Onset Date

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

- First day of month
- Informed consent date

- If year is after year of first dose date, impute as January 1 (YYYY-01-01)

- If year is before year of first dose date, impute as latest of:
  - January 1 (YYYY-01-01)
  - Informed consent date

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

3.7.1.3 Derivation of Treatment-Emergent Adverse Event Flag

- Sites are instructed to "split" AEs that are ongoing as of first dose date into two records: one that ends as of first dose date and one that begins on first dose date. Therefore in some cases, what may at first appear as an AE that begins on first dose date is really a continuation of an AE that began prior to first dose date. If all of the following criteria are true for any two AE records then the two records should be considered as one event:
  - The preferred term of the first AE = the preferred term of the second AE.
  - The start date of the first AE is before the date of first dose.
  - The end date of the first AE is one day before the date of first dose or equal to the date of first dose.
  - The start date of the second AE = the date of first dose.
  - The severity of the second AE ≤ the severity of the first AE. If collapsing AE for the purposes of reporting, use maximum severity.
  - The relatedness of both the first and the second AEs is "Not Related".
- After determining which records should be treated as a single event, the treatment-emergent flag can be determined as follows:
  - All AEs with an onset date on or after the first dose date and no later than 30 days after the last dose date.
    - The 30-day rule will not always apply for subjects in crossover. In such cases any event with an onset date on or after the start of brigatinib
  - In the case of missing onset date, impute using the rules defined in Section 3.7.1.2 then determine if imputed date is on or after first dose date.

3.7.1.4 Treatment-Related Adverse Event Flag

All AEs for which relationship to study treatment is classified as "Possibly Related", "Probably Related", or "Definitely Related" will be classified as treatment related.

If relationship to study treatment is missing, then the relationship will be imputed as treatment related.

3.7.1.5 Calculation Specifications for Treatment-Emergent Adverse Events

Time to onset is defined as the time interval from the first dose date until the onset date of a treatment-emergent AE, or first occurrence of any event in a group of AEs. In case of a missing onset date, a date imputed with the rules in Section 3.7.1.2 will be used. Time to onset will be computed with the following formula: time to onset= AE onset date – first dose date + 1.

Time to initial onset of an individual treatment-emergent AE is defined as the time interval from the first dose date until the earliest onset date of the treatment-emergent AEs of the same nature. In case of a missing onset date, a date imputed with the rules in Section 3.7.1.2 will be used. Time to initial onset will be computed with the following formula: time to initial onset = earliest AE onset date – first dose date + 1.

Duration of an adverse event, or group of AEs, is the sum of all days on or after randomization on which the event in question is reported. Each day will only be counted once in the case of overlapping intervals.

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

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

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

3.7.1.6 Specifications for the Analyses Outputs

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

3.7.1.7 Special Categories of Adverse Events

In order to best characterize the safety experience of brigatinib and crizotinib in this study additional analyses will be performed in which preferred terms are grouped according to potential safety signals of interest. These lists will be maintained by the ARIAD coding group in the Data Management department and will be updated whenever MedDRA is updated.

The special category for pulmonary adverse events will include the preferred terms "PNEUMONITIS" and "INTERSTITIAL LUNG DISEASE". Early Onset Pulmonary Events (EOPEs) are defined as treatment-emergent AEs with these preferred terms with onset within the first 14 days relative to first exposure to randomized treatment. Late onset pulmonary events occurring 15 days or more after start of therapy will also be tabulated.

Some categories utilize Standardized MedDRA Queries (SMQs) to retrieve the conditions of interest, while others use Modified MedDRA Queries based on SMQs (MMQs) or Customized MedDRA Queries (CMQs). The current list of categories to be analyzed is provided in Appendix 1.

Each special category will consist of a list of Preferred Terms, and any such terms in the AE database will be included in the category. A single preferred term can be included in more than one special category.

3.7.2 Laboratory and Vital Signs Data

Laboratory and vital signs data in standard units will be summarized using summary statistics and graphically at baseline and at each cycle for which adequate data are available. Laboratory and blood pressures data will also be graded according to the NCI CTCAE, v.4.03 when applicable. Change from baseline to the worst on-study result will be summarized using the changes from baseline values in standard units. Change from baseline to the worst on-study result will also be summarized by shift in CTCAE grade for selected laboratories and blood pressures. Changes in testosterone and insulin will be summarized by cross-tabulations of baseline (low, normal, high) and the highest/lowest value on study (low, normal, high).

Laboratory data received without reference ranges will have the values imputed using standard reference ranges (Kratz, et al, 2004). Results received as collected at the site will not include the imputed normal ranges. Standardized results, included on each observation of laboratory data, will include the imputed ranges and will be used for all analyses unless stated otherwise.

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

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

3.7.3 QT interval corrected (Fridericia) (QTcF) Analysis

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

3.7.4 Extent of Exposure

Exposure to brigatinib and crizotinib will be summarized using the following measures:

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

Time on treatment will be defined as the time interval from the first dose date to the last dosing date and computed with the following formula:

Time (days) on treatment= last non-zero dose date – first dose date + 1

Dose intensity will be calculated with the following formula: Dose intensity = Total cumulative dose/ Time (days) on study treatment. Relative dose intensity will be defined as the proportion of the planned dose received by subjects. In treatment Arm A the daily planned dose will be 90 mg in the 7-day lead-in period and 180 mg from day 8 onward. In treatment Arm B the daily planned dose will be crizotinib 250 mg twice daily (BID) orally.

Relative dose intensity will be calculated as follows:

Relative dose intensity = Total cumulative dose administered /

Total dose planned x 100%.

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

Total person years = Time (days) on study treatment / 365.25.

The total person years in an analysis population will be the sum of the total person years of all the subjects in this population.

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

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

Study drug administration records that indicate that treatment was received on a given date but for which a specific quantity of drug cannot be determined will be treated in the following manner:

- The date will be treated as a day on which drug was taken for calculations such as total days exposed
- It will be considered as a day of reduced dose in the dose modification analyses
- No contribution will be made to the calculation of total cumulative dose administered or related analyses

4 REFERENCES

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

Cox, David R. Regression Models and Life-Tables. Journal of the Royal Statistical Society, Series B. 1972; 34 (2): 187–220.

DeMets DL, Lan KK. Interim analysis: The alpha spending function approach. Stat Med. 1994 Jul 15-30;13(13-14):1341-52.

Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eu J Cancer 2009;45:228-247.

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

FDA Guidance: "Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics Guidance for Industry" (April 2015)

ICH. Guidance for Industry: E9 Statistical Principles for Clinical Trials (September 1998) Kaplan EL and Meier P. Nonparametric Estimation from Incomplete observations. J Am Statistical Assoc. 1958;53:457-481.

Kratz A, Ferraro M, Sluss PM, Lewandrowski KB. Laboratory Reference Values. N Engl J Med 2004;351:1548-63.

NCI, Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03: June 14, 2010

5 APPENDICES

Appendix 1 Adverse Events in Special Categories

Selection of AEs for special categories will be performed through the use of Standardised MedDRA Queries based on SMQs, Modified MedDRA Queries (MMQs), or Customized MedDRA Queries (CMQs). The adverse events in special categories will be listed and updated as new events occur and/or MedDRA versions change over the course of the study.

The following special categories of adverse events will be analyzed:

- Early Onset Pulmonary Events (EOPEs) and Later Onset Pneumonitis Events
- Bradycardia Events
- Hypertension Events
- Gastrointestinal (GI) Events
- Pancreatic Events
- Elevated Insulin/Hyperglycemia Events
- Hepatic Events
- Creatine Phosphokinase (CPK) Elevation
- Anemia and Lymphopenia
- Peripheral Neuropathy
- Skin and Subcutaneous Events
- Vision Impairment Events
- QT Interval Prolongation

Appendix 2 Details for Sensitivity Analyses of the Primary Endpoint

Table A3-1 Sensitivity 1 - Progression and Censoring for the Analysis of PFS

#Rule	Situation	Date of progression or censoring	Outcome
1.	Missing or incomplete baseline tumor assessment	Day 1 (randomization date)	Censored
2.	No valid post-exposure assessment	Day 1 (randomization date)	Censored
3.	No measurable disease at baseline	Date of new lesion(s) or substantial worsening in non-target disease Date of last evaluable progression-free radiographic assessment	Progressed Censored
4.	a) Death, PD, or non-PD observed > 126 days after last valid RECIST assessment	Date of last evaluable progression-free radiographic assessment prior to the 126-day interval	Censored
	b) Death or PD <= 126 days after last valid RECIST assessment	Earliest of date of death and date of 1st PD	Progressed
5.	No progression or death	Date of last radiological assessment of measured lesions	Censored
6.	Death before first PD assessment or between evaluable assessment visits	Date of death	Progressed
7.	Death after one missed radiographic assessment	Handled according to rule #4	
8.	Death after two or more missed radiographic assessments	Handled according to rule #4	
9.	Treatment discontinuation prior to documented disease progression or death – **Added in this analysis**	Date of last administration of randomized study drug	Censored
10.	New anticancer treatment started prior to PFS event	Date of last dose of treatment to which subject was randomized	Censored
11.	Cancer-related surgery prior to documented disease progression	Date of last evaluable progression-free radiographic assessment prior to surgery	Censored
12.	Disease progression documented between scheduled visits	Date of progression as reported by BIRC	Progressed
13.	Disease progression preceded by 1 missed follow-up disease assessment	Handled according to rule #4	
14.	Disease progression preceded by 2 or more consecutive missed follow-up disease assessments	Handled according to rule #4	

All censoring rules in the primary analysis other than the specific change above remain in effect as specified for that analysis.

Table A3-2 Sensitivity 2 - Progression and Censoring for the Analysis of PFS

#Rule	Situation	Date of progression or censoring	Outcome
1.	Missing or incomplete baseline tumor assessment	Day 1 (randomization date)	Censored
2.	No valid post-exposure assessment	Day 1 (randomization date)	Censored
3.	No measurable disease at baseline	Date of new lesion(s) or substantial worsening in non-target disease Date of last evaluable progression-free radiographic assessment	Progressed Censored
4.	a) Death, PD, or non-PD observed > 126 days after last valid RECIST assessment - **Modified**	**Earliest of date of death and date of 1st PD**	**Progressed**
	b) Death or PD <= 126 days after last valid RECIST assessment	Earliest of date of death and date of 1st PD	Progressed
5.	No progression or death	Date of last radiological assessment of measured lesions	Censored
6.	Death before first PD assessment or between evaluable assessment visits	Date of death	Progressed
7.	Death after one missed radiographic assessment	**Handled according to rule #4**	
8.	Death after two or more missed radiographic assessments	**Handled according to rule #4**	
9.	New anticancer treatment started prior to PFS event- **Modified**	**Ignored – treat as censored or progressed as determined by rule #4**	
10.	Cancer-related surgery prior to documented disease progression- **Modified**	**Ignored – treat as censored or progressed as determined by rule #4**	
11.	Disease progression documented between scheduled visits	Date of progression as reported by BIRC	Progressed
12.	Disease progression preceded by 1 missed follow-up disease assessment	**Handled according to rule #4**	
13.	Disease progression preceded by 2 or more consecutive missed follow-up disease assessments	**Handled according to rule #4**	

The changes in rules 4, 9, and 10 are the only specific change in this algorithm. However, the change to rule 4 carries over to the other bolded items in the table.

Appendix 3 Anticancer Therapies Treated as Potential Censoring Events

Concomitant anticancer therapies will be identified by medical review, and a list of therapy terms will be created to select all radiotherapy treatments.

The following are examples of therapies observed in brigatinib studies:

1. "PLEURADESIS"
2. "PLEURODESIS"
3. "BIFRONTAL CRANIOTOMY FOR TUMOR AND ICH REMOVAL"
4. "BIFRONTAL CRANIOTOMY FOR TUMOR AND ICH"
5. "BRAIN SURGEY"
6. "TALC PLEURODESIS"
7. "RESECTION"
8. "EXCISION OF LESION"
9. "WEDGE RESECTION OF SEGMENT 9 AND 1 OF THE RIGHT LUNG"

Additional potential censoring events will be selected from concomitant medications records using the following selection criterion:

> From concomitant medications dataset, any medication with ATC level 2 classification of "Antineoplastic agents" (except the following treatment: "FLOUROURACIL 5% TOPICAL CREAM")

All the identified therapies will be treated as potential censoring events in the PFS event algorithm and associated endpoints. Data from concomitant medications will be combined with other therapies and the start date of the first such therapy for each subject will be identified as the potential censoring event.

Appendix 4 Radiotherapies Treated as Potential PFS or Censoring Events

Radiotherapies will be identified by medical review, and a list of therapy terms will be created to select all radiotherapy treatments.

The following are examples of radiotherapies observed in previous brigatinib studies:

10. "GAMMAKNIFE SURGERY"
11. "HOLOCRANEAL RADIOTHERAPY"
12. "IRRADIATION OF BRAIN METASTASIS SRS"
13. "IRRADIATION SRS"
14. "PALLIATIVE RADIOTHERAPY"
15. "RADIATION TREATMENT FOR LEFT SHOULDER BONE METASTASIS"
16. "RADIOTHERAPY"
17. "RADIOTHERAPY - CYBERKNIFE (21 GY DOSE)TO TREAT THE BRAIN METASTASIS"
18. "RADIOTHERAPY 10GY/1FRACTION"
19. "RADIOTHERAPY 30GY IN 10 FRACTIONS"

Determination of whether the identified therapies will be treated as PFS events or censoring events will depend on whether the supporting data indicate that the treatment was directed at iCNS metastases. Therapies recorded as being directed at iCNS metastases will be treated as potential PFS events in the primary analysis. Radiotherapies reported as being directed at the brain will be included in the iCNS category. All other tumor-directed radiotherapies will be treated as potential censoring events.

It is possible that a subject could receive radiotherapy that is directed at both iCNS and another area of the body on the same day. In such an event the iCNS radiotherapy will take precedence and that date will be treated as a potential PFS event.

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	04-Jun-2020 17:25 UTC