Statistical Methods

Detailed methodology for summary and statistical analyses of the data collected in this trial will be documented in a Statistical Analysis Plan (SAP), which will be maintained by Pfizer.

Sample Size Determination

The primary objective of this study is to demonstrate that lorlatinib (Arm A) is superior to crizotinib (Arm B) in prolonging PFS by BICR assessment per RECIST v1.1. A key secondary objective of the study is to demonstrate that lorlatinib is superior to crizotinib in prolonging OS.

The study is designed to test H_0: $HR_{PFS} \geq 1$ vs. H_A: $HR_{PFS} < 1$, where HR_{PFS} is the hazard ratio (Arm A/Arm B) of PFS.

Approximately 280 patients (140 in each arm) were to be randomized using a 1:1 ratio stratified by presence of brain metastases and ethnic origin. As of 28 February, 2019 enrollment was closed with 296 patients randomized. One hundred seventy seven (177) PFS events will be required to have at least 90% power to detect a hazard ratio (HR) of 0.611 using a one-sided stratified log-rank test at a significance level of 0.025, and a 2-look group-sequential design with a Lan-DeMets (O'Brien-Fleming) α-spending function to determine the efficacy boundaries

The planned sample size was determined based on the assumption of a HR of 0.611 under the alternative hypothesis which, under an exponential model, assumes a median PFS of 11 months in the crizotinib arm and 18 months in the lorlatinib arm. CCI [redacted] The sample size further assumed a 15% drop-out rate within each treatment arm at 30 months and a non-uniform patient accrual over approximately 15 months and follow-up after the last patient is randomized of approximately 18 months

This sample size will also allow comparison of OS between the 2 treatment arms, provided that superiority of lorlatinib over crizotinib with respect to PFS has been demonstrated. If the true HR is 0.70 under the alternative hypothesis (under an exponential model, assumes median OS of 48 months on the crizotinib arm and 68.6 months on the lorlatinib arm), a total of 198 deaths will be required to have 70% power using a one-sided stratified log-rank test at a significance level of 0.025, and a 3-look group-sequential design with Lan-DeMets (O'Brien-Fleming) α-spending function to determine the efficacy boundaries.

These calculations further assumed a 15% drop-out rate for OS on either treatment arm at 120 months and a follow-up of approximately 110 months after the last patient is randomized.

		CYCLE 1 28 days			CYCLE 2 28 days	CYCLES ≥3 (28 days per cycle)	END of TREATMENT (EOT)/FOLLOW-UP		
Visit Identifier[a]	**Screening**[1] **(≤28 days prior to randomization)**	**Day 1**	**Day 8**	**Day 15**	**Day 1**	**Day 1**	**End of Treatment/ Withdrawal**[30]	**Post-Treatment Follow-Up**[31] **(every 4 weeks until PD)**	**Survival Follow-Up**[32] **(every 4 months up to 3 years, then every 6 months thereafter)**
Visit Time Window (days)	N/A		±1	±1	±2	±2	±3	±7	±7
ECOG performance status[7]	X	X			X	X	X		
Contraception check[8]	X	X	X	X	X	X	X	X	
Laboratory									
Hematology[9]	X	X(- 7 days)	X	X	X	X	X	X	
Blood Chemistry[10]	X	X (-7 days)	X	X	X	X	X	X	
HBV, HCV (if applicable)	X								
Lipids[11]	X	X(- 7 days)	X	X	X	X	X	X (at 30 days)	
Coagulation[12]	X	X (-7 Days)			X	X	X	X	
Urinalysis[13]	X	X(-7 days)			(X)	(X)	X	X	
Pregnancy test[14]	X	X			X	X	X		
Triplicate (12-lead) ECGs[15]	X	X (pre and 1-4 hs post dose for Arm A and pre-dose for Arm B)	X (1-4 hs post dose for Arm A only)	X(1-4 hs post dose for Arm A only)	X(1-4 hs post dose for Arm A only and 2-6 hs post dose for Arm B)	X (single reading, pre-dose, on Day 1 of every other cycle for both Arms)	X (Triplicate ECG for both Arms)		
LVEF assessment (Echocardiogram or MUGA scan)[16]	X	X (pre- dose- not to be repeated if normal and performed <2 weeks prior to randomization)				X (pre-dose, on Day 1 of every other cycle; a time window of 1 week prior to Day 1 is permitted)	X		

Exploratory Objectives:	Exploratory Endpoints:
[redacted]	[redacted]

This protocol will use an independent endpoint adjudication committee (BICR) to determine whether certain investigator-reported events meet the definition of disease-related efficacy endpoints, using predefined endpoint criteria.

3. STUDY DESIGN

This is a Phase 3, multinational, multicenter (at approximately 160 sites) , randomized, open-label, parallel 2-arm study in which approximately 280 patients with previously untreated advanced ALK-positive NSCLC were to be randomized 1:1 to receive lorlatinib monotherapy or crizotinib monotherapy according to the study design shown in Figure 1 (296 patients were actually randomized).

Patients will be stratified according to (also see Section 5.1.1):

- Presence of brain metastases (Yes vs No);
- Ethnic origin (Asian vs Non-Asian).

Crossover between treatment arms will not be permitted.

3.1. Study Treatment

This study was to randomize approximately 280 patients in a 1:1 ratio to receive:

- Arm A: Lorlatinib single agent;
- Arm B: Crizotinib single agent.

A cycle duration will be 4 weeks (28 days) and will always be considered 4 weeks irrespective of any dose delays/dosing interruptions or missed doses which may affect nominal days of each cycle.

Study treatment may continue until confirmed disease progression assessed by BICR, patient refusal, patient lost to follow-up, unacceptable toxicity, or the study is terminated by the sponsor, whichever comes first (see Section 6.4 Patient Withdrawal).

Patients who develop radiological disease progression confirmed by BICR assessment but are otherwise continuing to derive clinical benefit from study treatment will be eligible to continue with the treatment they have been assigned to, provided that the treating physician has determined that the benefit/risk for doing so is favorable. See Section 7.2 for details on expedited BICR assessment of disease progression.

Details of the study treatment forms and packaging and recommendations for dose modifications are included in the Study Treatment Section 5 of the protocol.

4. SUBJECT ELIGIBILITY CRITERIA

This study can fulfill its objectives only if appropriate patients are enrolled. The following eligibility criteria are designed to select patients for whom participation in the study is considered appropriate. All relevant medical and non-medical conditions should be taken into consideration when deciding whether a particular patient is suitable for this protocol.

Patient eligibility should be reviewed and documented by an appropriate member of the investigator's study team before patients are included in the study.

4.1. Inclusion Criteria

Patients must meet all of the following inclusion criteria to be eligible for enrollment into the study:

1. Diagnosis:

   a. Study Population: Patients with histologically or cytologically confirmed diagnosis of locally advanced [(Stage IIIB not amenable for multimodality treatment) or metastatic (Stage IV) by American Joint Committee on Cancer (AJCC) v 7.0] ALK-positive NSCLC where ALK status is determined by the FDA-approved (for use in US), CE (Conformité Européene) marked (for EU and other countries that accept CE marking), and PMDA(Pharmaceuticals and Medical Devices Agency)-approved (for use in Japan) Ventana ALK (D5F3) Companion Diagnostic (CDx) IHC test performed on the Ventana ULTRA or XT platforms (refer to Section 6.1.1.1 for any prescreening activity related to ALK determination);

   b. Tumor Requirements: At least 1 extracranial measurable target lesion per RECIST v. 1.1 that has not been previously irradiated. CNS metastases are allowed if **asymptomatic** and:

      1. Either untreated and not currently requiring corticosteroid treatment, or on a stable or decreasing dose of ≤10 mg QD prednisone or equivalent; or

For samples being collected and shipped, detailed collection, processing, storage, and shipment instructions and contact information will be provided to the investigator site prior to initiation of the study.

7.1. Tumor Response Assessments

Tumor assessments will include all known or suspected disease sites. Imaging will include chest, abdomen, and pelvis CT or MRI scans and brain MRI. CNS imaging using MRI (unless contraindicated) is required at baseline in all patients and at every tumor assessment. The response of measurable intracranial disease (ie, lesions ≥5 mm) will be assessed by a modified version of RECIST v1.1,[33] see Appendix 3. Tumor assessments are to be done at screening, then repeated every 8 weeks (±1 week) starting from randomization while on treatment and during follow-up until PD. Bone scans will be performed (or bone MRI if preferred by Investigator) at baseline and on study only if bony metastases are suspected. Patients with positive results on the bone scans (or bone MRI) will have these repeated every 16 weeks (±1 week).

Tumor assessments must continue until documented disease progression has been determined by BICR. If a patient continues treatment based on the investigator's judgement of clinical benefit in spite of disease progression- confimed by BICR- (eg, intracranial lesions are not in progression) tumor assessment must continue to be performed every 8 weeks ±1 week (16 weeks ±1 week for Bone Scan/MRI if applicable).

For patients discontinuing treatment for reasons other than progression of disease, tumor assessments must be evaluated for tumor response until PD, regardless of new anti-cancer therapy.

Both target and non-target lesions are to be followed using the same modality used at baseline and interval unless clinically indicated. Tumor assessments should be repeated after at least 4 weeks to confirm response, and at the End of Treatment visit if more than 8 weeks have passed since the last evaluation.

7.2. Expedited Blinded Independent Central Review for Disease Progression

An expedited BICR review will be performed for investigator-assessed disease progression. Upon investigator-assessed disease progression, all radiographic images collected for a patient from baseline onwards will be submitted to the BICR for expedited review (see the Study Manual for process details and maximum allowable time). Every effort should be made to keep the patient on study treatment and have all the assessments performed as per Schedule of Activities until the BICR has completed the radiographic image review, unless contraindicated by the investigator. Once the patient has documented PD by BICR, patients should be discontinued from study treatment.

However, if according to the Investigator's clinical judgment, a patient with evidence of PD is still experiencing clinical benefit, the patient may be eligible for continued treatment with the assigned drug after discussion between the Investigator and Pfizer. The Investigator's judgment should be based on the overall benefit/risk assessment, like the intracranial effect, and the patient's clinical condition, including performance status, clinical symptoms, adverse events and laboratory data. In this specific circumstance (eg, intracranial lesions are not in

progression), the tumor assessments will continue to be performed every 8 weeks (every 16 weeks for bone) (±1 week) until treatment stop or as clinically indicated.

A request for an expedited BICR review may also be triggered by Pfizer upon notification of cases that have been reported as PD by Parexel during their timepoint review, but not submitted for global expedited review yet since not identified as PD cases by the investigator.

7.3. Pregnancy Testing

For female patients of childbearing potential, a serum or urine pregnancy test, with sensitivity of at least 25 mIU/mL will be performed on 2 occasions prior to starting administration of study treatment, once at the start of screening and once at the baseline visit immediately before starting the investigational product administration. Following a negative serum pregnancy test result at screening, appropriate contraception must be commenced and another negative serum or urine pregnancy test result will then be required at the baseline visit before the patient may receive the investigational product. Urine pregnancy tests will also be routinely repeated at every treatment cycle during the active treatment period, at the end of study treatment, and additionally whenever 1 menstrual cycle is missed or when potential pregnancy is otherwise suspected. In the case of a positive confirmed pregnancy, the patient will be withdrawn from administration of investigational product but may remain in the study.

Additional pregnancy tests may also be undertaken if requested by Institutional Review Board (IRBs)/ECs or if required by local regulations.

7.4. Safety Assessments

Safety assessments will include collection of AEs, SAEs, vital signs and physical examination, 12 lead ECGs, echocardiogram/MUGA, laboratory assessments, including pregnancy tests and verification of concomitant treatments. Ophthalmologic examinations will be performed in all patients. Assessment for mood and suicidal ideation and behavior will also be performed (see Schedule of Activities).

7.4.1. Adverse Events

Assessment of AEs will include the type, incidence, severity (graded by the National Cancer Institute Common Terminology Criteria for Adverse Events [NCI CTCAE] version 4.03) timing, seriousness, and relatedness.

Adverse events that occur during the study, including baseline signs and symptoms, will be recorded on the AE CRF page.

7.4.2. Laboratory Safety Assessment

Hematology, blood chemistry, and urinalysis will be collected at the time points described in the Schedule of Activities table and analyzed at local laboratories. They may also be performed when clinically indicated and relevant results reported in the CRF as ‘unplanned visits’. The required laboratory tests are listed in Appendix 4. Local laboratory certification(s) and reference ranges should be provided to the sponsor prior to study patient screening activity.

7.4.3. Vital Signs and Physical Examinations

Patients will have a physical examination to include major body systems, body weight, blood pressure, pulse rate, assessment of ECOG performance status, and height (height will be measured at screening only) at the time points described in the Schedule of Activities. Blood pressure and pulse rate should be taken with the patient in the seated position after the patient has been sitting quietly for at least 5 minutes.

7.4.4. (12-Lead) Electrocardiograms

A triplicate 12-lead (with a 10-second rhythm strip) tracing will be used for all ECG assessments, except for ECG at Day 1 of Cycle ≥3, which will be single readings according to Schedule of Activities.

All patients require a triplicate ECG measurement at screening. On treatment ECGs will be performed as outlined in the SOA table. At each time point which requires a triplicate reading per SoA, 3 consecutive 12 lead ECGs will be performed approximately 2 minutes apart (or within 10 minutes, whichever is appropriate) to determine mean QTc (average of triplicates). CCI [redacted] Clinically significant findings seen on subsequent ECGs should be recorded as adverse events. In case of QTc >500 msec (ie, CTCAE Grade >2), ECG must be reviewed by qualified personnel at the site as soon as the finding is made, including verifying that the machine reading is accurate and that the Fridericia correction formula is applied. If the manual reading verifies a rate corrected QTc of >500 msec, repeat ECG should be immediately performed at least two times approximately 2 to 4 minutes apart.

An electronic reading of prolonged QTc must be confirmed by manual reading. Prior to conclusion that an episode of prolongation of the QTc is due to study drug, thorough consideration should be given to potential precipitating factors (eg, change in patient clinical condition, effect of concurrent medication, electrolyte disturbance) and possible evaluation by a specialist. If QTc reverts to less than 500 msec, and in the judgment of investigator and sponsor is determined to be due to a cause other than study drug, treatment may be continued with regular ECG monitoring.

If a patient experiences PR interval prolongation >200 msec or second-degree or third-degree AV block, while on treatment with lorlatinib, refer to Section 5.5.1.2. If a patient experiences bradycardia while on crizotinib, refer to the general recommendations for crizotinib dose modification in Section 5.5.2.

If patient experiences a cardiac or neurologic AE (specifically syncope, dizziness, seizures, or stroke), triplicate ECGs should be obtained at time of the event. If the mean QTc is prolonged (>500 msec), or new or worsened AV block is noted, then ECGs should be re- evaluated by a qualified person at the institution for confirmation and repeated as clinically indicated. CCI [redacted]

CCI [redacted]

7.6. CSF Collection for Lorlatinib and Crizotinib Concentration Measurements

Diagnostic CSF analysis is mandatory for patients with suspected or confirmed leptomeningeal disease/carcinomatous meningitis (LMD/CM) not visualized on MRI (optional for the remaining patients). To be performed at baseline, and if clinically safe and feasible, further CSF cytology will be performed as clinically indicated to assess anti-tumor response and to determine lorlatinib or crizotinib concentrations.

If a patient is required to undergo a lumbar puncture while on trial, an additional ~5 mL sample of CSF should be collected for determining CCI [redacted] Crizotinib and its metabolite (Arm B) concentrations. If a CSF sample is collected, CCI [redacted] at approximately the same time. The CSF concentrations of crizotinib, CCI [redacted] other potential metabolites may be determined using a validated or non-validated method. Detailed collection and shipment procedures will be provided in the Study Manual.

CCI [redacted]

CCI

[redacted]

Patients must complete all EORTC QLQ-C30, EORTC QLQ-LC13 and EQ-5D-5L self-assessment questionnaires in the clinic and these cannot be taken home. All scheduled assessments of the EORTC QLQ-C30, EORTC QLQ-LC13, and EQ-5D-5L must be completed in the clinic prior to any other study or medical procedures.

EORTC QLQ-C30 and QLQ-LC13:

The EORTC QLQ-C30 (Version 3.0) is a published, validated, and self-administered PRO questionnaire.[34] The EORTC QLQ-C30 consists of 30 questions and includes 5 functional scales (physical, role, cognitive, emotional, and social); a global health status/global quality of life scale; 3 symptom scales (fatigue, pain, nausea and vomiting); and 6 single items that assess additional symptoms (dyspnea, appetite loss, sleep disturbance, constipation, and diarrhea) and financial impact. All scales and single item measures range in score from 0 to 100. Higher scores on the functional scales represent higher levels of functioning. Higher scores on the global health status/quality of life scale represent higher health status/quality of life. Higher scores on symptom scales/items represent a greater presence of symptoms.

The EORTC QLQ-LC13 is the lung cancer-specific module of the EORTC Quality of Life Questionnaire.[35] The EORTC QLQ-LC13 consists of 13 questions and includes 1 multi-item scale and 9 single items assessing symptoms (dyspnea, cough, haemoptysis, and site-specific pain), side effects (sore mouth, dysphagia, peripheral neuropathy, and alopecia), and pain medication use. Similar to the EORTC QLQ-C30, higher scores are reflective of a greater presence of symptoms.

The EORTC QLQ-C30 and EORTC QLQ-LC13 will be administered as noted on the SOA table. Both the EORTC QLQ-C30 and QLQ-LC13 modules require about 15 minutes to complete.

EQ-5D-5L:

The EuroQol EQ-5D-5L is a patient-completed questionnaire designed to assess health status in terms of a single index value or utility score.[36,37] There are 2 components to the EuroQol EQ-5D-5L: a descriptive system in which individuals rate their level of problems (none, slight, moderate, severe, extreme/unable) in 5 areas (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a Visual Analogue Scale (VAS) in which patients rate their overall health status from 0 (worst imaginable) to 100 (best imaginable). Published weights are available that allow for the creation of a single summary score.

The EuroQol EQ-5D-5L questionnaire will be administered as noted on the SOA table. The amount of time for a patient to complete the EQ-5D-5L is estimated to be about 2 minutes.

7.10. Assessment of Mood

An assessment of mood will be administered to patients via the Beck Depression Inventory-II scale at the time points described in the Schedule of Activities table. This is a 21- item self-report scale, with each item rated by patients on a 4 point scale (ranging from 0-3). The scale includes items capturing mood (loss of pleasure, sadness, irritability), suicidal ideation, and cognitive signs (punitive thoughts, self- criticism, self-dislike, pessimism poor concentration) as well as somatic signs (appetite, sleep, fatigue, libido).

Safety Event	Recorded on the CRF	Reported on the CT SAE Report Form to Pfizer Safety Within 24 Hours of Awareness
SAE	All	All
Non-serious AE	All	None
Exposure to the investigational product under study during pregnancy or breastfeeding, and occupational exposure	All (regardless of whether associated with an AE), **except occupational exposure**	Exposure during pregnancy, exposure via breastfeeding, occupational exposure (regardless of whether associated with an AE)

All observed or volunteered events regardless of treatment group or suspected causal relationship to the investigational product(s) will be reported as described in the following paragraphs.

Events listed in the table above that require reporting to Pfizer Safety on the CT SAE Report Form within 24 hours of awareness of the event by the investigator **are to be reported regardless of whether the event is determined by the investigator to be related to an investigational product under study**. In particular, if the SAE is fatal or life-threatening, notification to Pfizer Safety must be made immediately, irrespective of the extent of available event information. This time frame also applies to additional new (follow-up) information on previously forwarded reports. In the rare situation that the investigator does not become immediately aware of the occurrence of an event, the investigator must report the event within 24 hours after learning of it and document the time of his/her first awareness of the event.

For each event, the investigator must pursue and obtain adequate information both to determine the outcome and to assess whether it meets the criteria for classification as an SAE (see the Serious Adverse Events section below). In addition, the investigator may be requested by Pfizer Safety to obtain specific follow-up information in an expedited fashion. This information is more detailed than that recorded on the CRF. In general, this will include a description of the event in sufficient detail to allow for a complete medical assessment of the case and independent determination of possible causality. Any information relevant to the event, such as concomitant medications and illnesses, must be provided. In the case of a subject death, a summary of available autopsy findings must be submitted as soon as possible to Pfizer Safety. Any pertinent additional information must be reported on the CT SAE Report Form; additional source documents (eg, medical records, CRF, laboratory data) are to be sent to Pfizer Safety **ONLY** upon request.

As part of ongoing safety reviews conducted by the sponsor, any non-serious AE that is determined by the sponsor to be serious will be reported by the sponsor as an SAE. To assist in the determination of case seriousness, further information may be requested from the investigator to provide clarity and understanding of the event in the context of the clinical study.

If a patient begins a new anticancer therapy, SAEs occurring during the above-indicated active collection period must still be reported to Pfizer Safety irrespective of any intervening treatment.

8.1.4.2. Recording Non-Serious AEs and SAEs on the CRF

During the active collection period, both non-serious AEs and SAEs are recorded on the CRF.

Follow-up by the investigator may be required until the event or its sequelae resolve or stabilize at a level acceptable to the investigator, and Pfizer concurs with that assessment.

If a patient begins a new anticancer therapy, the recording period for non-serious AEs ends at the time the new treatment is started; however, SAEs must continue to be recorded on the CRF during the above-indicated active collection period.

8.1.5. Causality Assessment

The investigator's assessment of causality must be provided for all AEs (serious and non-serious); the investigator must record the causal relationship on the CRF, and report such an assessment in accordance with the SAE reporting requirements, if applicable. An investigator's causality assessment is the determination of whether there exists a reasonable possibility that the investigational product caused or contributed to an AE; generally the facts (evidence) or arguments to suggest a causal relationship should be provided. If the investigator does not know whether or not the investigational product caused the event, then the event will be handled as "related to investigational product" for reporting purposes, as defined by the sponsor. If the investigator's causality assessment is "unknown but not related" to investigational product, this should be clearly documented on study records.

In addition, if the investigator determines that an SAE is associated with study procedures, the investigator must record this causal relationship in the source documents and CRF, and report such an assessment in the dedicated section of the CT SAE Report Form and in accordance with the SAE reporting requirements.

8.1.6. Sponsor's Reporting Requirements to Regulatory Authorities

AE reporting, including suspected unexpected serious adverse reactions, will be carried out in accordance with applicable local regulations.

8.2. Definitions

8.2.1. Adverse Events

An AE is any untoward medical occurrence in a study subject administered a product or medical device; the event need not necessarily have a causal relationship with the treatment or usage. Examples of AEs include, but are not limited to:

- Abnormal test findings;
- Clinically significant signs and symptoms;

- Changes in physical examination findings;
- Hypersensitivity;
- Drug abuse;
- Drug dependency.

Additionally, AEs may include signs and symptoms resulting from:

- Drug overdose;
- Drug withdrawal;
- Drug misuse;
- Drug interactions;
- Extravasation;
- Exposure during pregnancy (EDP);
- Exposure via breastfeeding;
- Medication error;
- Occupational exposure.

Worsening of signs and symptoms of the malignancy under study should be recorded as AEs in the appropriate section of the CRF. Disease progression assessed by measurement of malignant lesions on radiographs or other methods should not be reported as AEs.

8.2.2. Abnormal Test Findings

Abnormal objective test findings should be recorded as AEs when any of the following conditions are met:

- Test result is associated with accompanying symptoms; and/or
- Test result requires additional diagnostic testing or medical/surgical intervention; and/or
- Test result leads to a change in study dosing (outside of any protocol-specified dose adjustments) or discontinuation from the study, significant additional concomitant drug treatment, or other therapy; and/or
- Test result is considered to be an AE by the investigator or sponsor.

Hospitalization does not include the following:

- Rehabilitation facilities;
- Hospice facilities;
- Respite care (eg, caregiver relief);
- Skilled nursing facilities;
- Nursing homes;
- Same-day surgeries (as outpatient/same-day/ambulatory procedures).

Hospitalization or prolongation of hospitalization in the absence of a precipitating clinical AE is not in itself an SAE. Examples include:

- Admission for treatment of a preexisting condition not associated with the development of a new AE or with a worsening of the preexisting condition (eg, for workup of a persistent pretreatment laboratory abnormality);
- Social admission (eg, subject has no place to sleep);
- Administrative admission (eg, for yearly physical examination);
- Protocol-specified admission during a study (eg, for a procedure required by the study protocol);
- Optional admission not associated with a precipitating clinical AE (eg, for elective cosmetic surgery);
- Hospitalization for observation without a medical AE;
- Preplanned treatments or surgical procedures. These should be noted in the baseline documentation for the entire protocol and/or for the individual subject;
- Admission exclusively for the administration of blood products.

Diagnostic and therapeutic noninvasive and invasive procedures, such as surgery, should not be reported as SAEs. However, the medical condition for which the procedure was performed should be reported if it meets the definition of an SAE. For example, an acute appendicitis that begins during the reporting period should be reported if the SAE requirements are met, and the resulting appendectomy should be recorded as treatment of the AE.

considered important medical events, even before all other possible causes of liver injury have been excluded.

The threshold of laboratory abnormalities for a potential DILI case depends on the subject's individual baseline values and underlying conditions. Subjects who present with the following laboratory abnormalities should be evaluated further as potential DILI (Hy's law) cases to definitively determine the etiology of the abnormal laboratory values:

- Subjects with AST/ALT and TBili baseline values within the normal range who subsequently present with AST OR ALT values >3 × ULN AND a TBili value >2 × ULN with no evidence of hemolysis and an alkaline phosphatase value <2 × ULN or not available;

- For subjects with baseline AST **OR** ALT **OR** TBili values above the ULN, the following threshold values are used in the definition mentioned above, as needed, depending on which values are above the ULN at baseline:

    - Preexisting AST or ALT baseline values above the normal range: AST or ALT values >2 times the baseline values AND >3 × ULN; or >8 × ULN (whichever is smaller).

    - Preexisting values of TBili above the normal range: TBili level increased from baseline value by an amount of at least 1 × ULN **or** if the value reaches >3 × ULN (whichever is smaller).

Rises in AST/ALT and TBili separated by more than a few weeks should be assessed individually based on clinical judgment; any case where uncertainty remains as to whether it represents a potential Hy's law case should be reviewed with the sponsor.

The subject should return to the investigator site and be evaluated as soon as possible, preferably within 48 hours from awareness of the abnormal results. This evaluation should include laboratory tests, detailed history, and physical assessment. The possibility of hepatic neoplasia (primary or secondary) should be considered.

In addition to repeating measurements of AST and ALT and TBili, laboratory tests should include albumin, creatine kinase (CK), direct and indirect bilirubin, gamma-glutamyl transferase (GGT), prothrombin time (PT)/international normalized ratio (INR), total bile acids, alkaline phosphatase and acetaminophen drug and/or protein adduct levels. Consideration should also be given to drawing a separate tube of clotted blood and an anticoagulated tube of blood for further testing, as needed, for further contemporaneous analyses at the time of the recognized initial abnormalities to determine etiology. A detailed history, including relevant information, such as review of ethanol, acetaminophen (either by itself or as a coformulated product in prescription or over-the-counter medications), recreational drug, supplement (herbal) use and consumption, family history, sexual history, travel history, history of contact with a jaundiced person, surgery, blood transfusion, history of liver or allergic disease, and potential occupational exposure to chemicals, should be

Sensitivity analyses of PFS by BICR assessment will also be performed as described in the statistical analysis plan. Additionally, a Cox regression model, stratified for baseline stratification factors, will be used to explore the potential influences of the other factors on the primary PFS endpoint.

9.3.2. Analysis of Secondary Endpoints

All analyses will be performed using the FA set. The analysis of PFS will be repeated based on the Investigator's assessment.

The analyses of other tumor-related endpoints will be based on the BICR assessment, with the exception of OR that will be based both on BICR and on the investigator's assessment.

9.3.2.1. Overall Survival

Overall Survival (OS) is defined as the time from date of randomization to date of death due to any cause. Patients last known to be alive will be censored at date of last contact.

OS will be hierarchically tested for significance at the time of PFS interim/final analysis, provided the primary endpoint, PFS, is statistically significant at that analysis. In this case, OS will also be tested at 70% of the OS events (provided the first interim analysis of OS is not statistically significant) and at the OS final analysis (provided both interim analyses of OS are not statistically significant). A stratified log-rank test (one sided; using the same stratification factors used for PFS analysis) will be used at the interim and/or final analyses with the overall significance level preserved at 0.025 (one sided). OS time associated with each treatment arm will be summarized using the Kaplan- Meier method and displayed graphically where appropriate. CIs for the 25th, 50th, and 75th percentiles will be reported. The Cox proportional hazards model will be fitted to compute the treatment HRs and the corresponding 95% CIs.

Sensitivity analyses to adjust for the influence of subsequent therapy on the OS may be conducted, and further details will be specified in the statistical analysis plan.

9.3.2.2. Objective Response

Objective Response (OR) is defined as a complete response (CR) or partial response (PR) per RECIST version 1.1 (see Appendix 3) recorded from randomization until disease progression or start of new anti-cancer therapy. Both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. A patient will be considered to have achieved an OR if the patient has a sustained CR or PR according to RECIST version 1.1 definitions. Otherwise, the patient will be considered as a non-responder in the ORR analysis. Additionally, patients with inadequate data for tumor assessment (eg, no baseline assessment or no follow-up assessments) will be considered as non-responders in the ORR analysis.

The ORR on each treatment arm will be estimated by dividing the number of patients with OR (CR or PR) by the number of patients randomized to the respective treatment arm. The corresponding exact 2-sided 95% CIs will be provided. OR comparison between the 2 treatment arms will be assessed using Cochran-Mantel-Haenszel (CMH) test using the same stratification factors as for the PFS analysis.

In addition, the Best Overall Response (BOR) for each patient will be summarized by treatment arm.

9.3.2.3. Intracranial Objective Response

Intracranial Objective Response (IC-OR) is defined as above for OR but it is only based on intracranial disease in the subset of patients with at least 1 intracranial lesion.

The IC-OR will be summarized similar to how OR will be summarized as described above. If data permits, IC-OR will also be summarized in the subset of patients with at least 1 measurable intracranial lesion.

9.3.2.4. Intracranial Time to Progression

Intracranial Time to Progression (TTP) is defined as the time from randomization to the date of the first documentation of objective progression of intracranial disease, based on either new brain metastases or progression of existing brain metastases.

IC-TTP will be calculated for the FA set and for subgroups of patients with and without brain metastases at baseline. For the subgroup of patients without brain metastases at baseline only the new brain metastases will be considered events.

Differences between treatment arms will be assessed by the stratified log-rank test. IC-TTP associated with each treatment arm will be summarized using the Kaplan -Meier method and displayed graphically where appropriate. CIs for the 25th, 50th, and 75th percentiles will be reported. The Cox proportional hazards model will be fitted to compute the treatment HRs and the corresponding 95% CIs. Intracranial progression events may also be analyzed by cumulative incidence functions.

9.3.2.5. Duration of Response

Duration of Response (DR) is defined, for patients with an OR per RECIST version 1.1, as the time from the first documentation of objective tumor response (CR or PR) to the first documentation of objective tumor progression or death due to any cause, whichever occurs first. Censoring rules for DR will follow those described above for PFS.

DR will be summarized by treatment arm using Kaplan-Meier methods and displayed graphically, where appropriate. The median DR and 95% CI for the median will be provided for each treatment arm.

9.3.2.6. Intracranial Duration of Response

The IC-DR, based on BICR, will be summarized similar to DR as described above in the subset of the FA population with an IC-BOR of CR or PR, as the time from the first documentation of intracranial objective response (CR or PR) per RECIST 1.1 based on BICR to the date of first documentation of intracranial objective progression of disease (PD) or death due to any cause. If IC-OR will be summarized in the subset of patients with at least 1 measurable intracranial lesion, a corresponding IC-DR will be provided.

9.3.2.7. Time to Tumor Response

Time to Tumor Response (TTR) is defined, for patients with a confirmed OR, as the time from the date of randomization to the first documentation of objective response (CR or PR) which is subsequently confirmed.

TTR will be summarized using simple descriptive statistics (mean, SD (standard deviation), min, max, 25th, 50th, and 75th percentiles).

9.3.2.8. Intracranial Time to Tumor Response

Intracranial TTR based on BICR is defined, for patients with a confirmed IC-OR, as the time from the date of randomization to the first documentation of intracranial objective response (CR or PR) which is subsequently confirmed.

IC-TTR will be summarized using simple descriptive statistics (mean, SD, median, min, max, Q1, Q3).

9.3.2.9. Progression-Free Survival 2

PFS2 is defined as the time from randomization to the date of progression of disease on first subsequent systemic anti-cancer therapy, or death from any cause, whichever occurs first. If no date of disease progression on first subsequent systemic anti-cancer therapy is available, patients will be censored at date of last contact. A sensitivity analysis will be conducted including the date of discontinuation from first subsequent systemic anti-cancer therapy as an event.

Differences between treatment arms will be assessed by the stratified log-rank test. PFS2 time associated with each treatment arm will be summarized using the Kaplan-Meier method and displayed graphically where appropriate. CIs for the 25th, 50th, and 75th percentiles will be reported. The Cox proportional hazards model will be fitted to compute the treatment HRs and the corresponding 95% CIs.

9.3.2.10. Patient-Reported Outcomes

The EORTC QLQ-C30, EORTC QLQ LC13, and EQ-5D-5L will be scored according to their respective user guides/scoring manuals.[42,39]

For each treatment arm and at each time point, the number and percentage of patients who complete the EORTC QLQ-C30, EORTC QLQ-LC13, and EQ-5D-5L will be summarized in a table, as will the reasons for non-completion of these measures.

Patient-reported HRQoL, disease/treatment-related symptoms of lung cancer, and health status will be assessed. Summary statistics (mean (and SD), median, range, and 95% CI) of absolute scores will be reported for the items and subscales of the EORTC QLQ C30, the items and subscales of the EORTC QLQ-LC13, and the EQ-5D-5L VAS scale. The mean change of absolute scores from baseline (and 95% CI) will also be assessed. Line charts depicting the means and mean changes of items and subscales over time will be provided for each treatment arm.

The number and proportion of patients who improved, worsened, or remained stable for the symptom and functional domains, global QOL and single items of the EORTC QLQ-C30 and EORTC QLQ-LC13 will be summarized for each treatment arm.

Improvement, worsening, and stable categories will be determined over each cycle and summarized as an average by patient in the symptom scales, functioning scales, and in the global QOL scale. In the symptom scales, improvement is defined as a decrease of at least 10 points. In the functioning and global QOL scales improvement is defined as an increase of at least 10 points. In the symptom scales worsening is defined as an increase of at least 10 points. In the functioning scale and global QOL scales, worsening is defined as a decrease of at least 10 points. Global QOL, functioning scales, and symptom scales that have not improved nor worsened will be considered stable Osoba et al, established that a ≥10 point minimally important difference from baseline (ie, the first PRO measurement prior to initial treatment) on the scales of the EORTC QLQ-C30 would correlate with significant (moderate) change in disease symptoms and functioning.[41]

For the EQ-5D-5L health state profiles, the proportions of patients reporting “no”, “slight”, “moderate”, “severe”, or “extreme” problems will be reported at each time point.

Scales assessing pain in chest, dyspnea, and cough from the EORTC-QLQ-LC13 will be used to evaluate TTD of these symptoms. TTD in pain in chest, dyspnea, or cough individually and as a composite endpoint will be defined as the time from randomization to the first time the patient’s score shows a 10 point or greater increase after baseline in any of the 3 symptoms. Patients will be censored at the last time they completed a subscale assessment if they have not deteriorated.

TTD of the symptom subscales will be summarized using Kaplan-Meier methods. The estimated Kaplan-Meier plots will be provided, and the 1-sided log-rank test stratified by randomization stratification factors will be the primary method to compare the time to first deterioration between lorlatinib and crizotinib. The median TTD and 2-sided 95% CI for the median will also be provided based on the Brookmeyer-Crowley method.

Treatment arms will be evaluated based on the mean scores of the EORTC QLQ-C30 and EORTC QLQ-LC13 and the EQ-5D-5L utility and VAS scores obtained from longitudinal mixed-effects regression models that contain the baseline score as a covariate. Other exploratory PRO analyses may be performed subsequently as needed.

The frequency of patients experiencing TEAEs corresponding to system organ class and MedDRA preferred term will be reported. Adverse events will be summarized by worst NCI CTCAE v4.03 severity grade and relatedness to study treatment within each treatment arm.

In all summaries, emphasis will be on TEAEs, namely, those with initial onset or that worsen in severity after the first dose of study medication.

Adverse events leading to death or discontinuation of study treatment, events classified as NCI CTCAE v4.03 Grade ≥3, trial drug-related events, and serious adverse events will be considered with special attention. As appropriate, the difference in risk between treatment arms for adverse events of clinical interest may be further assessed as described in the SAP.

Detailed information collected for each adverse event will include a description of the event, duration, whether the adverse event was serious, intensity, relationship to study treatment, action taken, and clinical outcome.

9.5.2. Laboratory Abnormalities

Laboratory test results will be graded according to NCI CTCAE v4.03. The frequency of patients with laboratory test abnormalities will be summarized according to the worst grade for each laboratory test result.

For laboratory tests without an NCI CTCAE grade definition, results will be categorized as normal (within normal ranges), abnormal, or not done.

Shift tables will be provided to examine the distribution of laboratory abnormalities.

9.5.3. Electrocardiograms

All ECGs obtained during the study will be evaluated for safety. The triplicate data will be averaged, and all summary statistics and data presentations will use the triplicate averaged data. Any data obtained from ECGs repeated for safety reasons after the nominal time points will not be averaged along with the preceding triplicates.

QT intervals will be corrected for heart rate (QTc) using standard correction factors (ie, Fridericia's [default correction], Bazett's, and possibly a study-specific factor, as appropriate). Data will be summarized and listed for QT, HR, RR, PR, QRS, QTc.

Descriptive statistics (n, mean, median, standard deviation, minimum, and maximum) of corrected QT interval and other ECG parameters will be used to summarize absolute values and changes from baseline on treatment. Categorical analysis will be conducted for the maximum change from baseline in corrected QT, PR, and QRS and the maximum post-baseline corrected QT interval.

9.5.4. Left Ventricular Ejection Fraction

For patients with MUGA scans or echocardiograms, individual LVEF proportion (%) and its changes from baseline will be summarized by time point. The number of patients and the

Abbreviation	Term
	Events
CV	Coefficient of variation
DDI	Drug-drug interaction
DILI	Drug-Induced Liver Injury
DLT	dose-limiting toxicity
DMC	data monitoring committee
DNA	deoxyribonucleic acid
DR	Duration of Response
DU	dispensing unit
EC	ethics committee
ECG	electrocardiogram
EDTA	edetic acid (ethylenediaminetetraacetic acid)
EDP	exposure during pregnancy
EudraCT	European Clinical Trials Database
FA	Full Analysis
FDA	Food and Drug Administration (United States)
FFPE	formalin fixed paraffin embedded
FNA	fine needle aspiration
FSH	follicle-stimulating hormone
GCP	Good Clinical Practice
GGT	Gamma glutamyl transferase
HBV	hepatitis B virus
HCV	hepatitis C virus
HDL	High density lipoprotein
HDPE	High Density Polyethylene
HGFR	Hepatocyte Growth Factor Receptor
HIV	human immunodeficiency virus
HR	hazard ratio
HRQL	health-related quality of life
HRT	hormone replacement therapy
IA	interim analyses
IB	Investigator's Brochure
ICF	Informed consent form
ICH	International Conference on Harmonisation
IC-OR	Intracranial Objective Response
IC-DR	Intracranial Duration of Response
IC-TTR	Intracranial Time to Tumor Response
ID	identification
IHC	Immunohistochemistry
ILD	interstitial lung disease
IND	investigational new drug application
INR	international normalized ratio
IP	Investigational product

SCHEDULE OF ACTIVITIES (SOA)

The Schedule of Activities table provides an overview of the protocol visits and procedures. Refer to the STUDY PROCEDURES and ASSESSMENTS sections of the protocol for detailed information on each procedure and assessment required for compliance with the protocol.

The investigator may schedule visits (unplanned visits) in addition to those listed in the Schedule of Activities table in order to conduct evaluations or assessments required to protect the well-being of the patient.

		CYCLE 1 28 days			**CYCLE 2 28 days**	**CYCLES ≥3 (28 days per cycle)**	**END of TREATMENT (EOT)/FOLLOW-UP**		
Visit Identifier[a]	**Screening**[1] **(≤28 days prior to randomization)**	**Day 1**	**Day 8**	**Day 15**	**Day 1**	**Day 1**	**End of Treatment/ Withdrawal**[30]	**Post-Treatment Follow-Up**[31] **(every 4 weeks until PD)**	**Survival Follow-Up**[32] **(every 4 months up to 3 years, then every 6 months thereafter)**
Visit Time Window (days)	N/A		±1	±1	±2	±2	±3	±7	±7
Informed consent[2]	X								
Tumor history, including ALK status determination[3]	X								
Medical/oncological history (including prior medications and smoking history)	X								
Physical examination	X (full PE at screening only)	X			X	X	X	X	
Baseline signs and symptoms[4]		X							
Ophthalmologic examination[5]	X	To be repeated during the study in case of visual disturbances CTCAE grade increase, or as clinically indicated							
Height	X								
Weight[6]	X	X			X	X			
Vital signs[6]		X	X	X	X	X	X	X	

		CYCLE 1 28 days			CYCLE 2 28 days	CYCLES ≥3 (28 days per cycle)	END of TREATMENT (EOT)/FOLLOW-UP		
Visit Identifier[a]	**Screening**[1] **(≤28 days prior to randomization)**	**Day 1**	**Day 8**	**Day 15**	**Day 1**	**Day 1**	**End of Treatment/ Withdrawal**[30]	**Post-Treatment Follow-Up**[31] **(every 4 weeks until PD)**	**Survival Follow-Up**[32] **(every 4 months up to 3 years, then every 6 months thereafter)**
Visit Time Window (days)	N/A		±1	±1	±2	±2	±3	±7	±7
For France and Germany only: Transthoracic echocardiogram (TTE) to assess PAP and right heart function.(refer to Section 7.4.6	X	X (pre-dose- not to be repeated if normal and performed <2 weeks prior to randomization)				X (at least every 6 months during treatment)	X (to be performed only if patient received at least 3 months of study drug treatment and if the last previous TTE assessment was performed more than 4 weeks before)		
Randomization and Treatment									
Randomization[17]		X							
Arm A: lorlatinib		orally on a continuous QD dosing schedule							
Arm B: crizotinib		orally on a continuous BID dosing schedule							

2. Local treatment has been completed with full recovery from the acute effects of radiation therapy or surgery prior to randomization, and if corticosteroid treatment for these metastases has been withdrawn for at least 4 weeks with neurological stability; or

3. In case of leptomeningeal disease (LMD) or carcinomatous meningitis (CM) if visualized on magnetic resonance imaging (MRI), or if baseline CSF positive cytology is available.

c. Tissue Requirements: All patients must have an archival formalin fixed, paraffin embedded (FFPE) tissue specimen available and collected prior to randomization. If archived tissue is unavailable, then a mandatory *de novo* biopsy must be performed.

2. No prior systemic NSCLC treatment for advanced (Stage IIIB not amenable for multimodality treatment) or metastatic (Stage IV) disease, including molecularly targeted agents (eg, ALK TKIs), angiogenesis inhibitors, immunotherapy, or chemotherapy. Prior treatment for earlier Stages of the NSCLC only allowed if completed more than 12 months prior to randomization.

3. Eastern Cooperative Oncology Group Performance Status (ECOG PS) 0, 1, or 2.

4. Age ≥18 years (or ≥20 years as required by local regulation).

5. Adequate Bone Marrow Function, including:

a. Absolute Neutrophil Count (ANC) ≥ 1,500/mm^3 or ≥1.5 x 10^9/L;

b. Platelets ≥100,000/mm^3 or ≥100 x 10^9/L;

c. Hemoglobin ≥9 g/dL.

6. Adequate Pancreatic Function, including:

a. Serum total amylase ≤1.5 x upper limit of normal (ULN)*;

b. Serum lipase ≤1.5 x ULN.

*if total amylase >1.5 x ULN, but pancreatic amylase is within the ULN, then patient may be enrolled.

7. Adequate Renal Function, including:

a. Serum creatinine ≤1.5 x ULN or estimated creatinine clearance ≥60 mL/min as calculated using the method standard for the institution.

8. Adequate Liver Function, including:

CCI Additional triplicate ECGs may be performed as clinically indicated and relevant results reported in the CRF as 'unplanned visits'.

7.4.5. Echocardiograms/MUGA Scans

In order to monitor left ventricular ejection fraction (LVEF), an echocardiogram or MUGA will be performed at the time point described in the Schedule of Activities. The same method should be used at each time point.

7.4.6. Transthoracic echocardiogram (France and Germany only)

Two (2) patients receiving lorlatinib under compassionate use programme in France have been reported to have Pulmonary Arterial Hypertension (PAH), which have been diagnosed in absence of right heart catheterization. However, in one of these patients, PAH was a pre-existing condition and confounding factors were present in the other patient. In addition, 3 cases of Pulmonary Hypertension (PH) (1 not-related SAE, 1 not-related non-SAE, and 1 related non-SAE) were reported in study B7461001. No other cases were reported in approximately 900 patients treated with lorlatinib and no findings associated with pulmonary hypertension have been observed in lorlatinib nonclinical studies.

Although the current evidence does not indicate PH is a risk associated with lorlatinib treatment, investigators in Germany and France have been required, as requested by French (Agence National de Sécurité du Medicament et des produits de santé- ANSM) and German (Bundesinstitut für Arzneimittel und Medizinprodukte-BfArM) health authorities, to implement appropriate risk-reduction measures, including a trans thoracic echocardiogram (TTE) along the study treatment period for patients enrolled in these two countries.

In order to monitor pulmonary blood pressure, the condition of the heart valves, and ventricular motion, transthoracic echocardiogram (TTE) will be performed at the time points described in the Schedule of Activities. In this respect, the benefits/risks of including patients with cardiopulmonary disorders/comorbidities or risk factors should be carefully considered. In addition:

- A prompt TTE must be performed in the event of symptoms or signs suggesting PH, and all other explorations pursuant to PH diagnosis guidelines.
- Should PH occur during treatment, following multidisciplinary discussions involving a PH specialist (cardiologist or pulmonologist), consider dose reductions or even permanently discontinuing lorlatinib in the absence of haemodynamic and clinical recovery.
- Information must be provided to patients treated with lorlatinib on the new safety finding, with the recommendation to seek immediate medical attention in the event of signs suggesting PH (particularly dyspnoea and fatigue).

CCI

[redacted]

7.9. Patient-Reported Outcome Assessments

PROs will be assessed using the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)[34] and its corresponding module for lung cancer (QLQ-LC13)[35] and the EuroQol 5 dimension 5 level (EQ-5D-5L) questionnaire.[36,37] See PRO instruments included in Appendix 6.

7.11. Assessment of Suicidal Ideation and Behavior

To assess suicidal ideation behaviors, the Columbia Suicide Severity Rating Scale (C-SSRS)[37] will be administered to patients at the time points described in the Schedule of Activities table. The C-SSRS is a unique, simple and short method of assessing both behavior and ideation that tracks all suicidal events and provides a summary of suicidality. It assesses the lethality of attempts and other features of ideation (frequency, duration, controllability, reasons for ideation, and deterrents), all of which are significantly predictive of completed suicide.

7.12. Ophthalmologic Examinations

Ophthalmologic examinations for both right and left eyes will be performed at screening for all patients enrolled and should be performed by an ophthalmologist.

Ophthalmologic examinations should be repeated during the course of the study whenever a vision disorder AE is observed or CTCAE grade change occurs from a previous visit.

Best Corrected Visual Acuity and Refraction

Best corrected visual acuity will be assessed by using a standard wall or projection chart (Snellen, or another validated scale, provided that the result is converted to Snellen value) before implementing any procedures that can affect vision (eg, pupil dilation). The same chart will be used throughout the study for a specific patient, and the right eye should be tested first. The refractive error also will be determined. The ophthalmologist should ensure that patients are seated comfortably and that they do not move their head forward or backward during testing. Patients will be told that the chart contains only letters.

Biomicroscopy

Slit-lamp biomicroscopy will be performed without dilation of the pupil and should precede the administration of any pupil-dilating agent for ophthalmoscopy.

Any abnormalities (including severity) of the eyelids, conjunctivae, sclerae, corneas, anterior chambers, irises, or lens will be recorded.

Fundoscopy

Fundoscopy will be performed after pupillary dilation to examine the vitreous body, retina macula, retina non-macula (peripheral), optic nerve head, fundus and optic disc notching.

8. ADVERSE EVENT REPORTING

8.1. Requirements

The table below summarizes the requirements for recording safety events on the CRF and for reporting safety events on the Clinical Trial (CT) Serious Adverse Event (SAE) Report Form to Pfizer Safety. These requirements are delineated for 3 types of events: (1) SAEs; (2) non-serious adverse events (AEs); and (3) exposure to the investigational product under study during pregnancy or breastfeeding, and occupational exposure.

8.1.1. Additional Details on Recording Adverse Events on the CRF

All events detailed in the table above will be recorded on the AE page(s) of the CRF. It should be noted that the CT SAE Report Form for reporting of SAE information is not the same as the AE page of the CRF. When the same data are collected, the forms must be completed in a consistent manner. AEs should be recorded using concise medical terminology and the same AE term should be used on both the CRF and the CT SAE Report Form for reporting of SAE information.

8.1.2. Eliciting Adverse Event Information

The investigator is to record on the CRF all directly observed AEs and all AEs spontaneously reported by the study patient/legally acceptable representative. In addition, each study patient/legally acceptable representative will be questioned about the occurrence of AEs in a non-leading manner.

8.1.3. Withdrawal From the Study Due to Adverse Events (Also See Section 6.4 Patient Withdrawal)

Withdrawal due to AEs should be distinguished from withdrawal due to other causes, according to the definition of AE noted below, and recorded on the CRF.

When a subject withdraws from the study because of an SAE, the SAE must be recorded on the CRF and reported, as appropriate, on the CT SAE Report Form, in accordance with the Requirements section above.

8.1.4. Time Period for Collecting AE/SAE Information

The time period for actively eliciting and collecting AEs and SAEs ("active collection period") for each patient begins from the time the patient provides informed consent, which is obtained before the patient's participation in the study (ie, before undergoing any study-related procedure and/or receiving investigational product), through and including a minimum of 28 calendar days after the last administration of the investigational product.

For subjects who are screen failures, the active collection period ends when screen failure status is determined.

8.1.4.1. Reporting SAEs to Pfizer Safety

All SAEs occurring in a patient during the active collection period are reported to Pfizer Safety on the CT SAE Report Form.

SAEs occurring in a patient after the active collection period has ended are reported to Pfizer Safety if the investigator becomes aware of them; at a minimum, all SAEs that the investigator believes have at least a reasonable possibility of being related to investigational product must be reported to Pfizer Safety.

Follow up by the investigator continues throughout and after the active collection period and until the event or its sequelae resolve or stabilize at a level acceptable to the investigator, and Pfizer concurs with that assessment.

Merely repeating an abnormal test, in the absence of any of the above conditions, does not constitute an AE. Any abnormal test result that is determined to be an error does not require recording as an AE.

8.2.3. Serious Adverse Events

A serious adverse event is any untoward medical occurrence at any dose that:

- Results in death;
- Is life-threatening (immediate risk of death);
- Requires inpatient hospitalization or prolongation of existing hospitalization;
- Results in persistent or significant disability/incapacity (substantial disruption of the ability to conduct normal life functions);
- Results in congenital anomaly/birth defect.

Or that is considered to be:

- An important medical event.

Medical and scientific judgment is exercised in determining whether an event is an important medical event. An important medical event may not be immediately life-threatening and/or result in death or hospitalization. However, if it is determined that the event may jeopardize the subject or may require intervention to prevent one of the other AE outcomes, the important medical event should be reported as serious.

Examples of such events are intensive treatment in an emergency room or at home for allergic bronchospasm; blood dyscrasias or convulsions that do not result in hospitalization; or development of drug dependency or drug abuse.

Progression of the malignancy under study (including signs and symptoms of progression) should not be reported as an SAE unless the outcome is fatal within the active collection period. Hospitalization due to signs and symptoms of disease progression should not be reported as an SAE. If the malignancy has a fatal outcome during the study or within the active collection period, then the event leading to death must be recorded as an AE on the CRF, and as an SAE with Common Terminology Criteria for Adverse Events (CTCAE) Grade 5 (see the Severity Assessment section).

8.2.4. Hospitalization

Hospitalization is defined as any initial admission (even less than 24 hours) in a hospital or equivalent healthcare facility, or any prolongation of an existing admission. Admission also includes transfer within the hospital to an acute/intensive care unit (eg, from the psychiatric wing to a medical floor, medical floor to a coronary care unit, or neurological floor to a tuberculosis unit). An emergency room visit does not necessarily constitute a hospitalization; however, the event leading to the emergency room visit is assessed for medical importance.

8.3. Severity Assessment

Note the distinction between the severity and the seriousness of an AE. A severe event is not necessarily an SAE. For example, a headache may be severe (interferes significantly with the subject's usual function) but would not be classified as serious unless it met one of the criteria for SAEs, listed below.

GRADE	Clinical Description of Severity
0	No change from normal or reference range (This grade is not included in the Version 4.03 CTCAE document but may be used in certain circumstances.)
1	MILD adverse event
2	MODERATE adverse event
3	SEVERE adverse event
4	LIFE-THREATENING consequences; urgent intervention indicated
5	DEATH RELATED TO adverse event

8.4. Special Situations

8.4.1. Protocol-Specified Serious Adverse Events

There are no protocol-specified SAEs in this study with exception of those mentioned at Section 5.5.2.6 Severe Visual Loss for Crizotinib Patients. All SAEs will be reported to Pfizer Safety by the investigator as described in previous sections, and will be handled as SAEs in the safety database.

8.4.2. Potential Cases of Drug-Induced Liver Injury

Humans exposed to a drug who show no sign of liver injury (as determined by elevations in transaminases) are termed “tolerators,” while those who show transient liver injury, but adapt are termed “adaptors.” In some subjects, transaminase elevations are a harbinger of a more serious potential outcome. These subjects fail to adapt and therefore are "susceptible" to progressive and serious liver injury, commonly referred to as drug-induced liver injury (DILI). Subjects who experience a transaminase elevation above 3 times the upper limit of normal (× ULN) should be monitored more frequently to determine if they are an “adaptor” or are “susceptible.”

In the majority of DILI cases, elevations in aspartate aminotransferase (AST) and/or alanine aminotransferase (ALT) precede total bilirubin (TBili) elevations (>2 × ULN) by several days or weeks. The increase in TBili typically occurs while AST/ALT is/are still elevated above 3 × ULN (ie, AST/ALT and TBili values will be elevated within the same lab sample). In rare instances, by the time TBili elevations are detected, AST/ALT values might have decreased. This occurrence is still regarded as a potential DILI. Therefore, abnormal elevations in either AST OR ALT in addition to TBili that meet the criteria outlined below are considered potential DILI (assessed per Hy’s law criteria) cases and should always be

collected. Further testing for acute hepatitis A, B, C, D, and E infection and liver imaging (eg, biliary tract) may be warranted.

All cases demonstrated on repeat testing as meeting the laboratory criteria of AST/ALT and TBili elevation defined above should be considered potential DILI (Hy's law) cases if no other reason for the liver function tests (LFT) abnormalities has yet been found. **Such potential DILI (Hy's law) cases are to be reported as SAEs, irrespective of availability of all the results of the investigations performed to determine etiology of the LFT abnormalities.**

A potential DILI (Hy's law) case becomes a confirmed case only after all results of reasonable investigations have been received and have excluded an alternative etiology.

8.4.3. Exposure to the Investigational Product During Pregnancy or Breastfeeding, and Occupational Exposure

Exposure to the investigational product under study during pregnancy or breastfeeding and occupational exposure are reportable to Pfizer Safety within 24 hours of investigator awareness.

8.4.3.1. Exposure During Pregnancy

For both unapproved/unlicensed products and for marketed products, an exposure during pregnancy (EDP) occurs if:

- A female becomes, or is found to be, pregnant either while receiving or having been exposed (eg, because of treatment or environmental exposure) to the investigational product; or the female becomes or is found to be pregnant after discontinuing and/or being exposed to the investigational product.

- An example of environmental exposure would be a case involving direct contact with a Pfizer product in a pregnant woman (eg, a nurse reports that she is pregnant and has been exposed to chemotherapeutic products).

- A male has been exposed (eg, because of treatment or environmental exposure) to the investigational product prior to or around the time of conception and/or is exposed during his partner's pregnancy.

9.4. Analysis of Other Endpoints

CCI [redacted]

[redacted]

[redacted]

9.4.3. Biomarker Analysis for Secondary CCI [redacted] Endpoints

Biomarkers will be assessed separately for whole blood, serum, plasma, archival tumor tissue, and de novo tumor tissue biospecimens. In each case, summaries of baseline levels, changes from baseline (where appropriate), expression or genetic alterations will be reported. For continuous variables, summary statistics may include the mean, ratio to baseline, standard deviation, 25th, median, and 75th percentile, %CV and minimum/maximum levels of biomarker measures; for categorical variables, summary may include number and percentage, odds ratio, frequency statistics, as appropriate.

Data from biomarker assays may be analyzed using graphical methods and descriptive statistics such as Wilcoxon Signed Rank, Wilcoxon Rank Sum Test, Kaplan-Meier estimates of efficacy parameters (eg, PFS, OS) with biomarkers as a covariate, or Box plots (this list is not exhaustive.) The statistical approach will examine correlations of biomarker results with pharmacokinetic parameters and measures of anti-tumor efficacy.

CCI [redacted]

9.5. Safety Analysis

The SA set will be the primary population for safety evaluations. Summaries of AEs and other safety parameters will be provided, by treatment arm, as appropriate.

9.5.1. Adverse Events

Adverse events will be classified using the medical dictionary for regulatory activities (MedDRA) classification system. The severity of the toxicities will be graded according to the NCI CTCAE v4.03 whenever possible (http://ctep.info.nih.gov/reporting/ctc.html).

percentage whose maximum relative decrease from baseline in LVEF is greater than 20% will be calculated.

9.5.5. Ophthalmologic Data

For the ophthalmologic data, baseline is defined as the ophthalmologic exam performed at screening.

Best-corrected visual acuity examination results: changes from baseline will be summarized/listed for all patients from the SA set with a baseline and at least one post-baseline assessment.

Biomicroscopy (Slit Lamp) and Ophthalmoscopy (Fundoscopy) Exam Results: For baseline results, percentage of patients falling into each category of the examination status (normal, abnormal: mild, abnormal: moderate, abnormal: severe, not done) will be summarized for each structure by eye. For post-baseline results, percentage of patients falling into each category of the examination status (new findings/worsening of findings, no change, improvement of findings, not done, etc.) will be summarized for each structure by eye, for all patients with a baseline and at least 1 post-baseline assessment.

Additional summaries of ophthalmologic data will be considered as appropriate and described in the SAP.

9.5.6. Mood and Suicidal Ideation and Behavior Analyses

Changes across treatment will be described according to the Statistical Analysis Plan.

9.6. Interim Analysis

The interim analysis (IA) will be performed based on the FA set. Any safety evaluation at the time of the IA will be based on the SA set.

The goals of the IA are to allow early stopping of the study for efficacy and to assess the safety of lorlatinib. The IA will be performed with all patients randomized in the study.

The study is designed to have 1 IA and the final analysis based on the primary PFS endpoint as assessed by BICR. A Lan-DeMets (O'Brien-Fleming) α-spending function is used to determine the efficacy boundaries for PFS as shown in table below (Table 6).

Table 6. Stopping Boundaries for PFS

Analysis	Number of events (Information fraction)	Z scale	p-value (1-sided)
Interim	133 (75%)	Z <-2.337	P<0.01
Final	177 (100%)	Z <-2.012	P<0.022

The IA will be performed after approximately 133 PFS events (75% of the 177 events planned at the end of the study) If the value of the test statistic exceeds the efficacy boundary ($z < -2.337$, $p < 0.01$, if the interim analysis is performed after exactly 133 events), then the study is considered to have met its primary objective. If the results of the IA indicate serious

Abbreviation	Term
IRB	institutional review board
IRT	interactive response technology
IUD	intrauterine device
IVR	interactive voice response
IWR	interactive web response
LDL	Low density protein
LVEF	left ventricular ejection fraction
LFT	liver function test
LMD	leptomeningeal disease
LPLV	last patient last visit
MRI	Magnetic Resonance Imaging
MDZ	midazolam
MUGA	multigated acquisition
N/A	not applicable
NGS	next generation sequencing
NSCLC	non-small cell lung cancer
NTI	Narrow therapeutic index
OCT	optical coherence tomography
OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
PAH	Pulmonary Arterial Hypertension
PAP	Pulmonary Arterial Pressure
PCD	primary completion date
PD	Pharmacodynamics
PFS	Progression-Free Survival
P-gp	Permeability glycoprotein
PMDA	Pharmaceuticals and Medical Devices Agency
PK	Pharmacokinetics
PH	Pulmonary Hypertension
PI	Principal Investigator
PR	Partial response
PRO	Patient-Reported Outcome
PST	Potential Sight Threatening
PT	prothrombin time
QD	quaque die (every day)
RECIST	Response Evaluation Criteria In Solid Tumor
RNA	ribonucleic acid
RP2D	Recommended Phase 2 Dose
SAE	serious adverse event
SAP	statistical analysis plan
SCL	Supply Chain Lead
SD	Stable disease