NOVARTIS

Clinical Development

LDK378/Ceritinib/Zykadia®

LDK378A2407 / NCT02465528

A Phase II, Open Label, Multi-Center, Multi-Arm Study of Ceritinib in Patients with Advanced Solid Tumors and Hematological Malignancies Characterized by Genetic Abnormalities of Anaplastic Lymphoma Kinase (ALK)

Statistical Analysis Plan (SAP) – Final analysis

Author: Trial Statistician, [redacted]

Document type: SAP Documentation

Document status: Final

Release date: 20-Sep-2018

Number of pages: 63

Document History – Changes compared to previous final version of SAP

The CLDK378A2407 is a non-pivotal/non-registration study.

In July 2017, a first interim analysis (IA) for the Glioblastoma (GBM) cohort with the cut-off date of 07-Apr-2017 was performed. The results of the IA, in which a total of 12 GBM patients were evaluated, showed that ceritinib was generally well tolerated, and no new safety signals were identified. Unfortunately the primary endpoint of antitumor activity as measured by Disease Control Rate (DCR) at 16 weeks based on local investigator assessment was not met.

These results were presented to the LDK378 GPT in August 2017 and a decision was made to stop further enrollment in the GBM cohort as per protocol. The GPT also recommended to plan out a study close-out strategy based on these results. The study enrollment was halted in December 2017 based on the results of the IA and the significant challenges to identify eligible patients.

The SAP for the final close-out analysis was based on the SAP approved in CREDI on 16-Mar-2016. According to teams agreement, an abbreviated CSR will be written. Therefore, the analysis was simplified by reducing the number of analyses to those needed for an abbreviated CSR.

Compared to the planned analyses in protocol amendment 1, following changed:

- In Section 10.3, the protocol suggests to summarize concomitant medications and significant non-drug therapies. For the purpose of the abbreviated CSR, it was decided to only provide a listing.
- Section 10.5.3.3 states that shift tables will be generated for laboratory tests where CTCAE grades are not defined, using the low/normal/high/(low and high). It was decided to only provide shift tables for laboratory tests where CTCAE grades are defined. All laboratory data will be listed, with values flagged to show the CTCAE grades and the classifications relative to the laboratory normal ranges.

Table of contents

List of abbreviations

AE	Adverse event
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
bid	bis in diem/twice a day
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
[redacted]	[redacted]
DCR	Disease Control Rate
DMC	Data Monitoring Committee
DLT	Dose limiting toxicity
DOR	Duration of response
FAS	Full Analysis Set
eCRF	Electronic Case Report Form
IVR	Interactive Voice Response
IWR	Interactive Web Response
MedDRA	Medical Dictionary for Drug Regulatory Affairs
NCI	National Cancer Institute
o.d.	Once Daily
ORR	Objective response rate
OS	Overall Survival
PFS	Progression-Free Survival
[redacted]	[redacted]
PPS	Per-Protocol Set
PRO	Patient-reported Outcomes
qd	Qua'que di'e / once a day
QoL	Quality of Life
RAP	Report and Analysis Process
RECIST	Response Evaluation Criteria in Solid Tumors
SAP	Statistical Analysis Plan
SOC	System Organ Class
TFLs	Tables, Figures, Listings
TTR	Time to response
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes all planned analyses for the Clinical Study Report (CSR) of study LDK378A2407, a Phase II, Open Label, Multi-Center, Multi-Arm Study of Ceritinib in Patients with Advanced Solid Tumors and Hematological Malignancies Characterized by Genetic Abnormalities of Anaplastic Lymphoma Kinase (ALK).

All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock of the study data.

1.1 Study design

This is a phase II, open-label, multicenter, parallel multi-arm, POC clinical trial that will be assessing ceritinib as monotherapy in diverse populations of patients with various life threatening tumors other than NSCLC and hematological malignancies (see Table 1-1) characterized by ALK genetic alteration (and/or overexpression in some diseases).

Table 1-1 Tumor types of interest

Arm	Tumor types
T1	Anaplastic Large Cell Lymphoma (ALCL)
T2	Inflammatory Myofibroblastic Tumor (IMT)
T3	Glioblastoma (GBM)
T4	Inflammatory Breast cancer (IBC)
T5	Any other ALK+ tumor

Disease Control Rate (DCR) based on local investigator assessment is the primary objective of this study. For patients with solid tumors the assessment criteria will be RECIST 1.1. For GBM, RECIST 1.1 (for primary endpoint) and RANO [redacted] criteria will apply. For hematologic tumors, Cheson response criteria will apply. The DCR is defined as the proportion of ceritinib treated patients with complete response (CR), partial response (PR) or stable disease (SD) at 16 weeks from the start of the treatment. The DCR responses will need to be confirmed at least 4 weeks later by the same method. An adaptive design using a Bayesian Hierarchical model (BHM) is planned for the analysis of DCR.

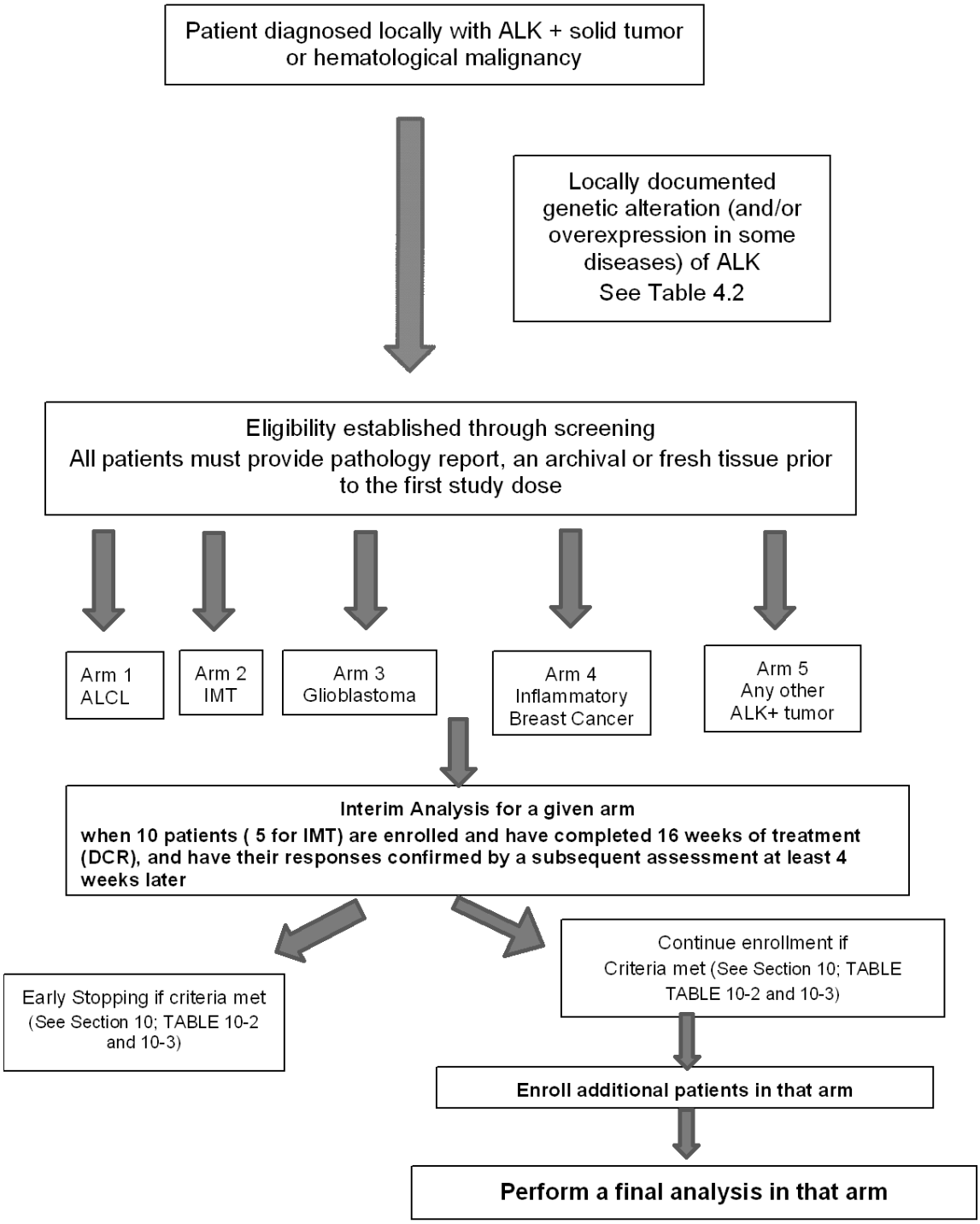
Interim analysis for each arm is planned for decision making (early stopping or continuation of recruitment) when a predefined number of patients (5 for IMT and 10 for all other arms) have completed 16 weeks of treatment with ceritinib and have their anti-tumor activity assessed and confirmed by a subsequent assessment at least 4 weeks later. An additional 10 patients may be enrolled for a given arm (including IMT) subsequent to the interim analysis

if the probability of the pre-defined clinically meaningful response for the arm is at least 20%. The final analysis for a given arm is planned to be performed when a predefined number of patients (15 for IMT and 20 for all other arms) have completed efficacy evaluation with confirmation of response (see Figure 1-1).

Due to adaptive nature of the proposed design the final sample size is not fixed. Approximately 106 patients are planned to be enrolled in this study in order to get 95 patients for the final analysis assuming an estimated attrition rate of 10%.

Given that the study was terminated based on the results of the first IA on GBM, the Bayesian hierarchical model will not be run for the final analysis.

Figure 1-1 Study design



Note: the references in this figure refer to sources in the protocol

1.2 Study objectives and endpoints

1.2.1 Primary objective

The primary objective is to assess the antitumor activity of ceritinib treatment as measured by Disease Control Rate (DCR) based on local investigator assessment.

Antitumor activity in each arm will be evaluated after 16 weeks of ceritinib treatment and will include responses of CR or PR or SD. The primary end point DCR at 16 weeks since the start of the treatment is considered clinically meaningful duration of drug activity in the selected patient population.

For patients with solid tumors the assessment criteria will be RECIST 1.1. For GBM, RECIST 1.1 (for primary endpoint) and RANO [redacted] criteria will apply. For hematologic tumors, Cheson response criteria will apply. DCR will include complete response (CR), partial response (PR) or stable disease (SD) at 16 weeks since the start of the treatment. The responses (CR or PR) will need to be confirmed at least 4 weeks later by the same method.

1.2.2 Secondary objectives

The key secondary objectives are to:

- assess the antitumor activity of ceritinib as measured by objective response rate (ORR), duration of response (DOR), time to response (TTR) determined by investigators
- assess the antitumor activity of ceritinib as measured by progression free survival (PFS) determined by investigators
- assess safety and tolerability:
  Hematology, biochemistry, urinalysis, coagulation, pregnancy test and hormones (males only); ECG; Performance status; Physical examination, Vital signs; Adverse events CTCAE 4.03 will be used to assess events.

Table 1-2 Objectives and related endpoints

Objective	Endpoint	Analysis
Primary		Refer to Section 1.2.1
To assess the antitumor activity of ceritinib as measured by DCR determined by investigators	DCR, defined as the proportion of patients with best overall response of CR, PR, or SD ≥ 16 weeks of ceritinib treatment. Responses (CR or PR) observed at week 16 must be confirmed by a subsequent assessment at least 4 weeks later.	
Secondary		Refer to Section 1.2.2
1. To assess the antitumor activity of ceritinib as measured by ORR determined by investigators	1. ORR, defined as the proportion of patients with a best overall response defined as CR or PR; (CR+PR)	
2. To assess the antitumor activity of ceritinib as measured by DOR,TTR and PFS determined by investigators	2. The following endpoints will be evaluated by investigator assessment per RECIST 1.1 and Cheson criteria: a. DOR, defined as the time from date of first documented CR or PR to date of first documented disease progression or death due to any cause b. TTR, defined as the time from date of the first dose to date of first documented response (CR or PR) c. PFS, defined as time from date of the first dose to date of first documented disease progression (assessed by investigators per or date of death due to any cause For solid tumor and GBM, RECIST 1.1 criteria will apply. For hematologic tumors, Cheson response criteria will apply.	
3. To assess the safety profile of ceritinib	3. ECG, Performance status, Vital signs, Physical examination; AEs (assessed by CTCAE 4.03), and laboratory (hematology, biochemistry, urinalysis, coagulation, pregnancy test and hormones (males only),	

Objective	Endpoint	Analysis
[redacted]		

2 Statistical methods

This section and its subsections will be be used to draft section 9.7 of the CSR after the analyses have been conducted.

In what follows, study drug refers to LDK378.

2.1 Data analysis general information

Data will be analyzed by Novartis Oncology Biostatistics and Statistical Programming personnel according to the data analysis plan in this SAP, which will be available in Appendix 16.1.9 of the Clinical Study Report (CSR). Important information is given in the following sections and details are provided, as applicable, in Section 4 from which Appendix 16.1.9 of the CSR will be extracted.

SAS® version 9.4 (or later version if available at time of database lock) will be used in all analyses.

Data from all patients who signed informed consent in centers that participate in this study will be used in the analysis. Data collected after withdrawal of informed consent will not be reported. Due to expected small size of enrollment at individual centers, no center effect will be assessed. Each analysis will use all data in the database up to the analysis cutoff date, determined prior to database lock.

Interim analysis for each arm will be performed for decision making (early stopping or continuation of recruitment) when a predefined number of patients (5 for IMT and 10 for all other arms) have completed 16 weeks of treatment with ceritinib and have their anti-tumor activity assessed. Responses observed at week 16 (CR or PR) must be confirmed by a subsequent assessment at least 4 weeks later. The reduced number of patients (5) in the IMT arm is based on the rare incidence of this tumor type. An additional 10 patients may be enrolled for a given arm (including IMT) subsequent to the interim analysis if the probability of the pre-defined clinically meaningful response for the arm is at least 20%.

For a specific arm, a decision will be made based on the calculated probability as given below:

- If the posterior probability of being clinically meaningful (Table 2-1) is less than 20% then recruitment will be stopped in that arm
- Otherwise, recruitment will be extended for at least 10 more patients in that arm

All available data will be used for analysis at the interim. However, a decision will be made only for arms with a minimum number of pre-specified patients (a minimum of 5 patients in IMT arm or a minimum of 10 patients for any of the other arm). Depending on enrollment and follow-up, there may be multiple interim analyses. Details of the interim analysis is given in Section 2.13.

The study will end approximately 38 weeks after the last patient started treatment with study medication or once at least 75% of patients have died or have been lost to follow-up or withdraws consent, or when the study is terminated early. The primary analysis for a given arm will occur once all patients in that arm, after initial accrual and the accrual following-up the

interim analysis, complete at least 16 weeks of treatment with ceritinib unless discontinued earlier; have their efficacy evaluation completed at week 16 and the responses (CR or PR) are confirmed by subsequent visit by a subsequent assessment at least 4 weeks later. The final analysis of study data will be conducted at the end of the study.

Section 2.2 provides further details regarding data to be included in the analyses.

Table 2-1 Type of tumors of interest with definition of being clinically meaningful

Disease Code	Arm	#Patients† at Interim	Not Clinically meaningful (C_1)	Clinically meaningful (C_2)
T1	Anaplastic large cell lymphoma (ALCL)	10	≤40%	≥50%
T2	Inflammatory myofibroblastic tumor (IMT)	5	≤20%	≥30%
T3	Glioblastoma	10	≤10%	≥20%
T4	Inflammatory breast cancer	10	≤10%	≥20%
T5	Any other ALK+ tumor	10	≤10%	≥20%
† Extend recruitment for at least 10 more patients in each arm according to IA outcome [p(clinically meaningful)>20%].				

General analysis conventions

Pooling of center: Unless specified otherwise, data from all study centers will be pooled for the analysis. Due to expected small size of centers, no center effect will be assessed.

Qualitative data (e.g., gender, race, etc.) will be summarized by means of contingency tables by tumor type; a missing category will be included as applicable. Percentages will be calculated using the number of patients in the relevant population or subgroup as the denominator.

Quantitative data (e.g., age, body weight, etc.) will be summarized by appropriate descriptive statistics (i.e. mean, standard deviation, median, minimum, and maximum) by tumor type.

2.1.1 General definitions

Study drug and study treatment both refer to LDK378 and will be used interchangeably.

Date of first/last administration of study drug

The date of first administration of study drug is defined as the first date when a non-zero dose of study drug was administered and recorded on the Dosage Administration Record (DAR) eCRF. The date of first administration of study drug will also be referred to as start date of study drug.

The date of last administration of study drug is defined as the last date when a non-zero dose of study drug was administered and recorded on DAR eCRF. This date will also be referred to as last date of study drug.

Study day

The study day for all assessments (e.g. tumor assessment, death, disease progression, tumor response, performance status, adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption) will be calculated using the start date of study drug as the origin. For assessments occurring after or on the start date of study drug, study day will be calculated as:

Study Day =Event date - start date of study drug + 1.

The first day of study drug is therefore study day 1.

For any assessment or events such as baseline disease characteristics or medical history (e.g., time since diagnosis of disease) occurring prior to the start of the study drug, study day will be negative and will be calculated as:

Study Day = Event date - start date of study drug.

Time unit

A year length is defined as 365.25 days. A month length is 30.4375 days (365.25/12). If duration is reported in months, duration in days will be divided by 30.4375. If duration is reported in years, duration in days will be divided by 365.25.

Baseline

For safety and efficacy evaluations, the last available assessment on or before the date of start of study treatment is defined as "baseline" assessment. If an assessment is planned to be performed prior to the first dose of study drug in the protocol and the assessment is performed on the same day as the first administration of study drug, it will be assumed that it was performed prior to study drug administration, if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 for baseline calculation.

Patients who start treatment and discontinue from the study on the same day may have 2 different sets of data collected on study day 1, one being reported to the cycle 1 day 1 visit, the other reported to the end of treatment (EOT) visit. Data reported at the EOT visit are not eligible for baseline selection.

If patients have no value as defined above, the baseline result will be missing.

On-treatment assessment/event and observation periods

The overall observation period will be divided into three mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent to the day before first administration of study treatment

2. ***on-treatment period***: from date of first administration of study treatment to 30 days after date of last actual administration of any study treatment (including start and stop date). For ongoing patients, on-treatment period is from date of first administration of study treatment to the data cut-off date (including start and cut-off date).
3. ***post-treatment period***: starting at day 31 after last dose of study drug

The safety summaries (tables, figures) include only assessments collected no later than 30 days after study drug discontinuation and assessments prior to the data cut-off date for on-going patients, unless otherwise specified. Specifically, data from the post-treatment period with the exception of deaths will not be included unless requested from Health Authorities or external committees.

For select items, shift tables or change from baseline summaries generated for laboratory, ECG, vital signs and change score generation may use data from pre-treatment period for baseline calculations.

All summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period, the so-called ***treatment-emergent*** AEs.

However, all safety data (including those from the post-treatment period) will be listed and those collected during the pre-treatment and post-treatment period are to be flagged.

Windows for multiple assessments

In order to summarize performance status/[redacted]/vital sign/other data collected over time (including unscheduled visits), the assessments will be time slotted. The following general rule will be applied in creating the assessment windows: If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If 2 assessments within a time window are equidistant from the target date, then the earlier of the 2 assessments will be used.. If multiple assessments on the same date then the the average will be used. Data from all assessments (scheduled and unscheduled), including multiple assessments, will be listed.

2.2 Analysis sets

A patient is considered to be enrolled into the study if he/or she has signed informed consent. Only a patient who has signed informed consent will be included in the analysis data sets.

Full Analysis Set

The Full Analysis Set (FAS) comprises all patients who receive at least one dose of Ceritinib. Unless otherwise specified the FAS will be the default analysis set used for all efficacy analyses, including the primary efficacy analysis.

Safety Set

The Safety Set includes all patients who received at least one dose of study medication.

Interim Analysis Set

The Interim Analysis Set comprises all patients who received ≥ 16 weeks of ceritinib treatment or discontinued early, and had their response confirmed by subsequent assessment at least 4 weeks later.

NOTE: The interim analysis set was not explicitely specified in the protocol. At the time of the first IA it was considered appropriate to perform the analysis of the primary endpoint based on the interim analysis set.

Any blood samples with missing blood collection date or time, or missing associated study drug dosing date or time will be excluded from PAS.

Patient Classification:

Patients may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific subject classification rules defined in Table 2-2.

Table 2-2 Subject classification based on protocol deviations and non-PD criteria

Analysis set	Protocol deviations leading to exclusion	Non protocol deviation leading to exclusion
FAS	No written inform consent	Not applicable
Safety set	No written inform consent	No dose of study medication

[redacted]

Withdrawal of Informed Consent

Any data collected in the clinical database after a subject withdraws informed consent from all further participation in the trial, will not be included in the analysis data sets. The date on which a patient withdraws full consent is recorded in the eCRF.

[redacted]

2.2.1 Subgroup of interest

Not applicable.

2.3 Patient disposition, demographics and other baseline characteristics

The FAS will be used for all patient demographic and baseline characteristic summaries and listings with the exception of screen failures.

Basic demographic and background data

All demographic and baseline disease characteristics data will be summarized and listed. Categorical data (e.g. gender, race, ethnicity, smoking history, WHO performance status, etc.) will be summarized by frequency count and percentages. Continuous data (e.g. age, weight, height, BMI, etc.) will be summarized by descriptive statistics (as defined in Section 2.1). BMI (kg/m2) will be calculated as weight[kg] / (height[m]2) using weight at Baseline.

Diagnosis and extent of cancer

Descriptive statistics and frequency counts and percentages will be tabulated, as appropriate, for diagnosis and extent of cancer based on the data collected on the electronic Case Report Form (eCRF) including primary site of cancer, tumor histology/cytology, histological grade, time (in months) since initial diagnosis of primary site, stage at initial diagnosis, time (in months) from initial diagnosis to first recurrence/progression, time (in months) since most recent relapse/progression, current extent of disease (metastatic sites), types of lesions (target and non-target lesions) at baseline, and disease burden at baseline for target lesion (based on the data collected on the RECIST eCRF page for solid tumor, on the RANO eCRF page for GBM, and on the Cheson response criteria eCRF page for hematologic tumors).

Medical history

Medical history and ongoing conditions, including cancer-related conditions and symptoms will be listed. Medical history and current medical conditions are coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology.

Prior anti-cancer therapy

Prior anti-neoplastic (anti-cancer) therapy will be listed in three separate categories: 1. medications, 2. radiotherapy, and 3. surgery.

The number and percentage of patients who received any prior anti-neoplastic medications, prior anti-neoplastic radiotherapy or prior anti-neoplastic surgery will be summarized.

Prior anti-neoplastic medications will be summarized by chemotherapy (medication) setting and other therapy (medication) setting.

Others

All data collected at baseline will be listed.

2.3.1 Patient disposition

The FAS will be used for the patient disposition summary tables and listings. The following will be tabulated:

- Number (%) of patients who are still on-treatment (based on the absence of the 'End of Treatment Phase Completion' page);
- Number (%) of patients who discontinued treatment (based on completion of the 'End of Treatment Phase Completion' page with date of discontinuation and reason of discontinuation/ 'Subject Status' entered);
- Primary reasons for study treatment discontinuation (based on discontinuation reasons entered under 'Subject Status' in the 'End of Treatment Phase Completion' page);

Protocol deviations

All protocol deviations will be listed.

Analysis sets

The number (%) of patients in each analysis set (defined in Section 2.2) will be summarized by tumor type.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

The Safety Set will be used for all medication data summaries and listings unless otherwise specified.

2.4.1 Study treatment / compliance

Duration of exposure, actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized by tumor type. Duration of exposure will be categorized into time intervals; frequency counts and percentages will be presented for the number(%) of subjects in each interval. The number (%) of subjects who have dose changes, reductions or interruptions, and the reasons, will be summarized by tumor type.

Subject level listings of all doses administered on treatment along with dose change reasons will be produced.

The safety set will be used for all summaries and listings of study treatment.

Dose exposure and intensity

Definitions of duration of exposure, cumulative dose, average daily dose, dose intensity (DI), relative dose intensity (RDI), as well as intermediate calculations, include:

- Duration of exposure (days): last date of study drug − first date of study drug + 1 (periods of interruption are not excluded)
- Cumulative dose (mg): total dose of study drug taken by a patient in the study
- Number of dosing days (days): duration of exposure − number of zero dose days
- Average daily dose (mg/day): cumulative dose (mg)/number of dosing days (days)
- DI (mg/day): cumulative dose (mg)/duration of exposure (days)
- RDI (%): 100 × [DI (mg/day)/planned dose (750 mg)]

Note: given the planned LDK378 dose of 750mg/day, the planned dose intensity can be calculated as:

PDI (mg/day) = cumulative planned dose (mg)/Duration of exposure (days),

where

cumulative planned dose (mg) = Protocol planned dose of 750 (mg)* Duration of exposure.

then, RDI (%) which is calculated as 100*DI/PDI can be simplified as shown above.

Dose reductions, interruptions or permanent discontinuations

The number of subjects who have dose reductions, permanent discontinuations or interruptions, and the reasons, will be summarized separately for each tumor type.

'Dose changed', 'Dose interrupted', and 'Dose permanently discontinued' fields from the Dosage Administration CRF pages (DAR) will be used to determine the dose reductions, dose interruptions, and permanent discontinuations, respectively.

The corresponding fields 'Reason for dose change/dose interrupted' and 'Reason for permanent discontinuation' will be used to summarize the reasons.

A dose change is either 'change in prescribed dose level' or 'dosing error' where actual dose administered/total daily dose is different from the prescribed dose.

For the purpose of summarizing interruptions and reasons, in case multiple entries for interruption that are entered on consecutive days with different reasons will be counted as separate interruptions. However, if the reason is the same in this block of entries, then it will be counted as one interruption.

Reduction: A dose change where the prescribed dose level is lower than the previous prescribed dose level or where the actual dose administered/total daily dose is lower than the calculated dose amount based on the prescribed dose. Therefore any dose change to correct a dosing error

will not be considered a dose reduction. Number of reductions will be derived programmatically based on the change and the direction of the change.

2.4.2 Prior, concomitant and post therapies

Prior anti-cancer therapy

The number and percentage of patients who received any prior anti-neoplastic medications, prior anti-neoplastic radiotherapy or prior anti-neoplastic surgery will be summarized by tumor type. Prior anti-neoplastic medications will be summarized by therapy type (e.g. chemotherapy, hormonal therapy etc.), setting (e.g. adjuvant, metastatic, etc.) and tumor type (arm). Summaries will also include total number of regimens. The medication therapy type of any combination therapy will be classified based on the following order: chemotherapy, biologic therapy, targeted therapy, hormonal therapy. For example, a combination therapy of chemotherapy and hormonal therapy will be classified as 'chemotherapy'. For radiotherapy, time since last radiotherapy, locations and setting of last therapy will be summarized. For prior surgery, time since last surgery and procedure will be summarized.

Separate listings will be produced for prior anti-neoplastic medications, radiotherapy, and surgery. Palliative antineoplastic radiotherapy administered during the treatment phase may also be summarized by tumor type.

Anti-neoplastic medications will be coded using the WHO Drug Dictionary (WHO-DD); anti-neoplastic surgery will be coded using MedDRA. Details regarding MedDRA and WHO-DD version will be included in the footnote in the tables/listings.

The above analyses will be performed using the FAS.

Concomitant medications

Concomitant therapies are defined as any medications (excluding study drug, prior antineoplastic treatments and blood transfusions), surgeries or procedures (including physical therapy) administered in the study and are recorded in the Prior and Concomitant Medications and the Surgical and Medical Procedures eCRF, respectively.

Concomitant medications will be coded using the WHO Drug Reference Listing (WHO DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (WHO ATC) classification system. Surgeries or procedures will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology.

All concomitant therapies will be listed. Any concomitant therapies starting and ending prior to the start of study drug or starting more than 30 days after the last date of study drug will be flagged in the listings.

2.5 Analysis of the primary objective

The primary objective is to assess the antitumor activity of ceritinib as measured by DCR determined by investigators in patients with ALK-positive ALCL, IMT, GBM, IBC or any other tumor type that is ALK positive, except from NSCLC. Patients must have received at least one line of prior systemic treatment for recurrent, locally advanced and/or metastatic disease.

2.5.1 Primary endpoint

The variable used to evaluate the primary objective is DCR associated with ceritinib treatment based on local investigator assessment. For patients with solid tumors or GMB the assessment criteria will be RECIST 1.1 and will include responses of CR or PR or SD ≥ 16 weeks after the first dose of the study drug. For hematologic tumors, Cheson response criteria will apply. Responses (CR or PR) should be confirmed by a subsequent assessment at least 4 weeks later.

In general, the definition of SD is applicable to patients with measurable disease only. If a patient without measurable disease is enrolled in the study, the non-CR/non-PD response for these patients will be considered equivalent to an SD response in endpoint determination.

2.5.2 Statistical hypothesis, model, and method of analysis

This study will contain 5 arms with ALK+ patients, an adaptive design will be used to assess activity of treatment in terms of response (DCR) rate within each arm and across each arm. The rules for how the trial adapts are pre-specified and are not based on ongoing data.

The design of the trial adapts to the data that are accumulated in the trial in such a way as to accommodate two possibilities:

a. DCR to ceritinib is specific to ALK+ inhibition and independent of the tumor site (similar DCR irrespective of tumor type);
b. the various tumor types have distinct DCR to ceritinib.

The pre-specified analysis is adaptive in the sense that when response depends on ALK+ inhibition and not on arm (scenario A), then it borrows from across the various arms and provide more precise estimate of DCR rates for those arms with similar response rates. In the other possibility (the various arms have distinct anti-tumor activity to ceritinib (Scenario B), the design allows little/no borrowing across tumor types. In this case the trial will be similar to traditional stratified analysis.

A hierarchical model (HM) will be used to analyze the binary data to facilitate the borrowing as specified above. Response rates (π_j) will be inferred for arm j (= 1,…,5). For each arm j, the number of responders follows a binomial distribution;

$$r_j \sim \text{Bin}(n_j, \pi_j)$$

We further let the parameters $\log (\pi_j/ (1 - \pi j))$ (logistic transformation) be either exchangeable with some of the arms, or non-exchangeable with any of them. Based on the number of strata in this trial, we allow for two exchangeability distributions, which, for example, accounts for the case where some arm shows no efficacy and some are promising.

Thus, for each arm j three possibilities arise, with respective probabilities $p_j = (p_{j1}, p_{j2}, p_{j3})$, as follows:

1. With probability p_{j1} the parameter θ_j follows a normal distribution with exchangeability parameters μ_1 and τ_1:

$$\theta_j \sim N(\mu_1,\tau_1^2)$$

2. With probability p_{j2}, θ_j follows a normal distribution with exchangeability parameters $\mu_1 < \mu_2$ and τ_2:

$$\theta_j \sim N(\mu_2,\tau_2^2)$$

3. With remaining probability $p_{j3} = 1-p_{j1}-p_{j2}$, θ_j follows a weakly-informative prior distribution

$$\theta_j \sim N(m_w,v_w)$$

For the detailed specifications of m_w, v_w, the a-priori weights p_j (j=1,…,J), and the prior distributions for μ_1, τ_1, μ_2, and τ_2, see (Appendix 5.4.1). At any given time of the trial, including at the end, posterior probabilities of the various parameters will be estimated using Markov chain Monte Carlo methods.

For a specific disease group, a Proof of Concept about treatment with ceritinib will be declared if both of the following conditions are met:

a. Observed DCR ≥ "Disease Control Rate" threshold (column for C_2 in Table 2-1)
b. Posterior probability of "not being clinically meaningful" (column for C_1 in Table 2-1) is less than 20%

2.5.3 Handling of missing values/censoring/discontinuations

Confirmed partial or complete responses or stable disease reported prior to any additional anticancer therapy will be considered as responses in the calculation of the DCR irrespective of the number of missed assessments before response.

For solid tumor, patients with a best overall response of 'Unknown' or 'Not Assessed' per RECIST v1.1 will be considered as non-responders and will be included in the denominator in estimating the DCR. For GMB or hematologic tumor, patients with unknown or missing response, or who are treated in the study but provide no information on response at the end of treatment will be treated as non-responders and will be included in the denominator when calculating DCR.

2.5.4 Supportive analyses

As a sensitivity analysis ORR and its 95% confidence interval will also be provided based on the exact binomial distribution (Clopper & Pearson, 1934).

2.6 Analysis of the key secondary objective

2.6.1 Efficacy

All secondary efficacy assessments (DOR, ORR, PFS, TTR) will be analyzed as per investigator assessment. Confirmation of response is required for all response endpoints, as per appropriate criteria (RECIST 1.1 for solid tumor and GMB, and Cheson for hematologic tumor).

All secondary analyses will be performed based on the FAS for each arm, unless otherwise specified.

No adjustment for multiple testing will be made.

The key secondary efficacy endpoints are:

- ORR
- DOR

- TTR
- PFS

The definitions and details on the derivation of the secondary endpoints are given in Section 14 (Appendix 2) of the LDK378A2407 protocol. Further details and rules needed for programmatic implementation of RECIST 1.1 guidelines and hematology criteria are provided in Section 5.1.

Overall Response Rate

Overall Response Rate (ORR) is based on local investigator assessment per RECIST 1.1 or Cheson criteria. For patients with solid tumors and GBM, the assessment criteria will be RECIST 1.1 and will include responses of CR and/or PR. For hematologic tumors, Cheson criteria will apply for evaluation of response. ORR is defined as the proportion of patients with best overall response of CR, PR. For ORR, estimate and its 95% CI will be provided.

Best overall response

The BOR will be assessed based on reported lesion responses at different evaluation time points. Both CR and PR must be confirmed by repeat assessments performed not less than 4 weeks after the criteria for response are first met. The next scheduled assessment may be used for purposes of confirmation of response.

BOR for each patient is determined from the sequence of overall (lesion) responses according to the following rules:

- CR = at least two determinations of CR at least 4 weeks apart before progression
- PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR)
- SD = at least one SD assessment (or better) > 6 weeks after start of study drug (and not qualifying for CR or PR).
- PD = progression ≤ 12 weeks after start of study drug (and not qualifying for CR, PR or SD).
- UNK = all other cases (i.e. not qualifying for confirmed CR or PR and without SD after more than 6 weeks or progression within the first 12 weeks)

Only tumor assessments performed before the start of any further anti-neoplastic therapies (i.e. any additional secondary anti-neoplastic therapy or anti-cancer surgery) will be considered in the assessment of BOR. If a patient receives any further anti-neoplastic therapy while on study, any subsequent assessments will be excluded from the BOR determination. Further anti-neoplastic therapies will be identified via protocol deviations or from the data collected on 'Anti-neoplastic therapies since last date of study drug' as appropriate. Clinical deterioration will not be considered as documented disease progression. Patients with BOR 'unknown' will be summarized by reason for having unknown status. The following reasons will be used:

- No valid post-baseline assessment
- All post-baseline assessments have overall response UNK

- New anti-neoplastic therapy started before first post-baseline assessment
- Stable disease (SD) too early (≤ 6 weeks after start date of study drug)
- PD too late (> 12 weeks after start date of study drug)

Special (and rare) cases where BOR is 'unknown' due to both early SD and late PD will be classified as 'SD too early'.

Duration of Response

Duration of Response (DOR) will be based on local investigator assessment per RECIST 1.1 for solid tumor, RANO and RECIST 1.1 for GBM, or Cheson criteria for hematologic tumors. Among patients with a confirmed response (PR or CR), DOR is defined as the time from first documented response (PR or CR) to the date of first documented disease progression or death due to any cause. If a patient has not had an event, DOR is censored at the date of last adequate tumor assessment. DOR will be listed by patient and may be described using Kaplan-Meier curves and relevant statistics if appropriate.

Time to response

Time to response (TTR) is defined as the time from the date of the first dose of LDK378 to first documented response (CR or PR, which must be confirmed subsequently) per appropriate critetia (RECIST 1.1 for solid tumor and GMB, and Cheson for hematologic tumor). TTR will be described using Kaplan-Meier methods and appropriate summary statistics. The TTR analysis will be conducted with censoring rules as described in the Appendix 5.1.1.

Progression-free survival

Progression-free survival (PFS) is defined as the time from the date of first dose of Ceritinib to the date of first documented disease progression per appropriate criteria (RECIST 1.1 for solid tumor and GMB, and Cheson for hematologic tumor) or death due to any cause.

A patient who has not progressed or died at the date of the analysis or when he/she receives any further anticancer therapy in the absence of disease progression will be censored at the time of the last adequate tumor evaluation before the earlier of the cut-off date or the anticancer therapy date. By default, if disease progression or death is documented after one single missing tumor evaluation, the actual event date of disease progression/death will be used for the PFS event date. If disease progression or death is documented after two or more missing tumor evaluations, the PFS time of these patients will be censored at the date of the last adequate tumor evaluation without PD.

If the study primary objective is not met (did not meet Proof of Concept criteria) for a given arm, Novartis may decide not to conduct some of the above secondary efficacy analyses in that arm but instead may choose to provide those endpoints in listings only.

Waterfall plot to depict anti-tumor activity

Waterfall graphs will be used to depict the anti-tumor activity. These plots will display the best percentage change from baseline in the sum of diameters of all target lesions for each patient with solid tumor (See Section 5.1.1 for details).

2.7 Analysis of secondary efficacy objective(s)

Not applicable.

2.8 Safety analyses

All safety analyses will be performed based on the Safety Set. All listings and tables will be presented by arm and for all patients within the arm.

The overall observation period will be divided into three mutually exclusive segments:

1. pre-treatment period: from day of patient's informed consent to the day before first dose of study medication

2. on-treatment period: from day of first dose of study medication to **30** days after last dose of study medication

3. post-treatment period: starting at day **31** after last dose of study medication.

2.8.1 Adverse events (AEs)

AE summaries will include all AEs occurring during on treatment period. All AEs collected in the AE (e)CRF page will be listed along with the information collected on those AEs e.g. AE relationship to study drug, AE outcome etc. AEs with start date outside of on-treatment period will be flagged in the listings.

AEs will be summarized by number and percentage of subjects having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding (using the latest version available prior to clinical database lock) by maximum severity (based on Common Terminology Criteria for Adverse Events [CTCAE] grades version 4.03), and relation to study drug. A subject with multiple occurrences of an AE will be counted only once in the respective AE category. A subject with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event.

In AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency.

AEs will be coded using MedDRA using the latest version available locket the time of analysis and will be graded using Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death. This grading system inherently places a value on the importance of an event, although there is not necessarily proportionality among grades (a grade 2 is not necessarily twice as bad as a grade 1). The CTCAE grade of 5 (death) is not used; rather, 'fatal' is collected as AE outcome and death information is also collected on a separate (e)CRF page.

The following AE summaries will be produced by tumor type:

- AEs regardless of study drug relationship
- AEs suspected to be study drug related
- On-treatment deaths, by primary system organ class and preferred term
- SAEs regardless of study drug relationship
- SAEs suspected to be study drug related
- AEs associated with discontinuation of study drug
- AEs requiring dose adjustment or study drug interruption
- AEs requiring significant additional therapy
- AEs excluding SAEs

All AEs regardless of study drug relationship will be summarized by tumor type and all patients.

2.8.1.1 Adverse events of special interest / grouping of AEs

NA for the abbreviated CSR for final analysis.

2.8.2 Deaths

Summary for on-treatment will be produced by tumor type, system organ class and preferred term. All deaths will be listed, post treatment deaths will be flagged.

2.8.3 Laboratory data

Data from all sources (central and local laboratories) will be combined. The summaries will include all laboratory assessments collected no later than 30 days after study drug discontinuation. All laboratory assessments will be listed and those collected later than 30 days after study drug discontinuation will be flagged in the listings.

Laboratory data will be classified (by Novartis Oncology Statistical Programming) into CTC grades according to the NCI CTCAE v4.03. For all reports, CTC grade is always obtained on the converted measurement in SI unit. Grade 5 will not be used.The CTC grade 0 will be assigned as below in different scenarios:

1. For lab parameters defined by criteria based on normal range only, a severity grade of 0 will be assigned when the value is within normal limits.
2. For lab parameters whose grade is defined by criteria based on normal range and absolute values (e.g. platelet count decrease). A severity grade of 0 will be assigned when the value is within normal limits.
3. For lab parameters whose grade is defined by criteria based on normal range and the change from baseline value, with no other associated clinical criteria such as concomitant medication (e.g. creatinine increased), the following will be applied. For the baseline grading and for the grading of post-baseline lab values with missing baseline grading, the grade will be derived using the criteria based only on the normal range as per CTCAE

v4.03. A severity grade of 0 will be assigned when the post-baseline value is ≤ ULN (for hyper) or ≥ LLN (for hypo).

Parameters for which a grading does not exist will be classified into low/normal/high group by means of laboratory normal ranges.

The following summaries will be produced for the hematology and biochemistry laboratory data (by laboratory parameter):

- Shift tables using CTC grades to compare baseline to the worst post-baseline value for laboratory parameters with CTC grades

The following lab parameters will be summarized:

- Hematology: absolute lymphocytes , absolute neutrophils , hemoglobin (anemia), WBC; platelet counts, absolute basophils, absolute eosinophils, absolute monocytes, RBC
- Biochemistry: alkaline phosphatase (APT), SGPT (ALT), SGOT (AST), total bilirubin, amylase, lipase, potassium (hyper and hypo), sodium (hyper and hypo), creatinine, glucose (fasting) (hyper and hypo), phosphate; albumin, calcium (at screening only corrected for albumin), magnesium, creatinine clearance, total bilirubin, blood urea nitrogen (BUN) or urea, GGT.
- For bi-directional parameters, both hyper and hypo summaries will be presented.
- Hormones (males only): Testosterone, LH, FSH, sex hormone binding globulin (SHBG).

The following laboratory parameters will be presented in listings and will not be summarized:

- Urinalysis: Macroscopic Panel (Dipstick) (Color, bilirubin, Blood, Glucose, Ketones, Leukocyte esterase, nitrite, pH, protein, specific gravity, urobilinogen).
- Microscopic panel (RBC, WBC, casts, crystals, bacteria, epithelial cells)
- Coagulation: INR

The following listings will be produced for the laboratory data for all laboratory parameters where CTC grades are defined:

- Listing of patients with laboratory abnormalities of CTC grade 3 or 4
- Listing of all laboratory data with values flagged to show the corresponding CTC grades and the classifications relative to the laboratory reference ranges.

Liver function tests (LFTs) of interest for LDK378 are total bilirubin (TBILI), ALT, AST and ALP. In what follows, AT refers to ALT or AST values. LFTs will be summarized as follows:

- Frequency counts and percentages of patients with worst post-baseline on-treatment values in the categories:
  - ALT > 3xULN, ALT > 5xULN, ALT > 8xULN, ALT > 10xULN, ALT > 20xULN
  - AST > 3xULN, AST > 5xULN, AST > 8xULN, AST > 10xULN, AST > 20xULN
  - AT > 3xULN, AT > 5xULN, AT > 8xULN, AT > 10xULN, AT > 20xULN
  - TBILI > 2xULN, TBILI > 3xULN
  - Concurrent AT > 3xULN and TBILI > 2xULN
  - Concurrent AT > 3xULN and TBILI > 2xULN and ALP < 2xULN
  - Concurrent AT > 3xULN and TBILI > 2xULN and ALP ≥ 2xULN

Concurrent measurements are those occurring on the same date.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

ECG data will be analyzed based on central laboratory reported results. The summaries will include all ECG assessments performed no later than 30 days after the last date of study drug. All ECG assessments will be listed, and those collected later than 30 days after study drug discontinuation will be flagged in the listing.

Selecting Primary QT Correction for Heart Rate

The analysis of QT data is complicated by the fact that the QT interval is highly correlated with heart rate. Because of this correlation, formulas are routinely used to obtain a corrected value, denoted QTc, which is independent of heart rate. This QTc interval is intended to represent the QT interval at a standardized heart rate. Several correction formulas have been proposed in the literature. For this analysis we will use QTcF as described below.

the QT interval corrected for heart rate by the Fridericia's formula, QTcF, is defined as

$$QTcF = \frac{QT}{\sqrt[3]{RR}},$$

where where RR=60/HR represents the RR interval of the ECG, in seconds.

Data will be summarized using QTcF.

ECG Summaries

The following analyses will be performed for each applicable ECG parameters (QT, ventricular rate – denoted as HR in what follows, and QTcF) as noted.

- Frequency counts and percentages of patients having notable ECG values according to the following categories:
  - QT parameter (QT, QTcF) increase from baseline >30 ms, >60 ms
  - Newly occurring post-baseline QT parameter > 450 ms, > 480 ms, > 500 ms
  - HR increase from baseline > 25% and value > 100 bpm
  - HR decrease from baseline > 25% and value < 50 bpm

  The denominator to calculate percentages for each category is the number of patients at risk for a specific category. For new abnormality post baseline, the denominator is the number of patients with both a baseline and a post-baseline evaluation and baseline not meeting the criterion. A newly occurring post-baseline ECG notable value is defined as a post-baseline value that meets the criterion post-baseline but did not meet the criterion at baseline.

All ECG assessments will be listed by patient and ECG parameter. In the listings, clinically notable values will also be flagged.

2.8.4.2 Vital signs

Vital sign assessments will be performed in order to characterize basic body function. The parameters collected are weight (kg), body temperature (°C), pulse rate (beats per minute), systolic and diastolic blood pressure (mmHg), and respiration rate (breaths per minute).

Clinically notable elevated values are defined as:

- Systolic BP: ≥ 160 mmHg and an increase ≥ 20 mmHg from baseline
- Diastolic BP: ≥ 100 mmHg and an increase ≥ 15 mmHg from baseline.
- Body temperature: ≥ 39.1°C
- Weight: increase from baseline of ≥ 10%
- Pulse rate: ≥ 120 bpm with increase from baseline of ≥15 bpm

Clinically notable below normal values are defined as:

- Systolic BP: ≤ 90 mmHg and a decrease ≥ 20 mmHg from baseline
- Diastolic BP: ≤ 50 mmHg and a decrease ≥ 15 mmHg from baseline
- Body temperature: ≤ 35°C
- Weight: decrease from baseline of ≥ 10%
- Pulse rate: ≤ 50 bpm with decrease from baseline of ≥ 15 bpm

All vital sign assessments will be listed by patient and vital sign parameter.

In the listings, clinically notable values will also be flagged.

2.8.4.3 ECOG performance status (WHO)

The ECOG performance assessment allows patients to be classified as to their functional impairment, the definition of scores in relation to their performance status is provided in Table 2-3, ranging from 0 (most active) to 5 (dead):

Table 2-3 ECOG performance scale

Score	Description
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
5	Dead

ECOG performance status at each time point will be listed.

2.10 Patient-reported outcomes

Not applicable.

2.12 Other Exploratory analyses

Not applicable

2.13 Interim analysis

Subjects will be continuously accrued and the data will be analyzed using a Bayesian hierarchical model. Interim analysis for an arm will be performed when a predefined number of patients (5 for IMT or T2 and 10 for all other tumor types) are enrolled and have completed 16 weeks of follow-up after the first dose of study drug and have their responses (CR or PR)

confirmed by a subsequent assessment at least 4 weeks later. At each of these analyses, the current (posterior) probability that the response rate for each of the tumor types is greater than "clinically meaningful (column for C_2 in Table 2-1)" response rate will be determined.

These probabilities will be used to adapt the design as follows:

1. Accrual to an arm that has at least predefined number of subjects will cease for futility if: it is very unlikely (probability <20%) that the response rate for the tumor is "clinically meaningful" (≥column for C_2 in Table 2-1)
2. Accrual to an arm will cease when its maximum planned sample size is reached. The results within an arm will be regarded as positive if the accrual for the arm reaches maximum and the conditions for positive conclusion are satisfied.

Table 2-4 Type of tumors of interest with definition of being clinically meaningful

Disease Code	Arm	#Patients† at Interim	Not Clinically meaningful (C_1)	Clinically meaningful (C_2)
T1	Anaplastic large cell lymphoma (ALCL)	10	≤40%	≥50%
T2	Inflammatory myofibroblastic tumor (IMT)	5	≤20%	≥30%
T3	Glioblastoma	10	≤10%	≥20%
T4	Inflammatory breast cancer	10	≤10%	≥20%
T5	Any other ALK+ tumor	10	≤10%	≥20%
† Extend recruitment for at least 10 more patients in each arm according to IA outcome [p(clinically meaningful)>20%].				

Note that if there are at least 5 patients of the same arm in "Any other ALK+ tumor, T5)" arm, then a separate arm will be opened for that specific tumor type leading to 5 different arms. However, interim analysis for decision making will not be performed until there are 10 patients in each of the arm: "new tumor type" and modified "Any other ALK+ tumor".

Since all patients will be treated with ceritinib in this open label study, it is not necessary to control access to interim results. Therefore, interim analysis will be performed by the study team to decide weather to terminate a given tumor type for futility. Note that recruitment will not be stopped while interim analysis is performed. This SAP decrribes the plan for interim analysis and no additional SAP that will be developed for the interim analysis.

3 Sample size calculation

The design of this study is adaptive in nature; hence, the final sample size is not fixed. A minimum of 10 and a maximum of 20 subjects in each arm T1, T3, T4 and T5 will receive treatment. For arm T2, the minimum and maximum sample size will be 5 and 15 respectively. Thus, the total sample size across all tumor types will be between 45 and 95. Assuming a 10% drop out rate, approximately 106 patients are planned to be enrolled in this study.

Subjects will be continuously accrued, and, the accumulated data will be analyzed at interim as described in Section 2.13. Based upon the results of any of these analyses, enrollment into one or more arms may be terminated.

The operating characteristics for this Bayesian Design are evaluated using simulation and are presented below. Simulation-based probability estimates (relative frequencies) of futility at interim, positive results at final analysis and average sample size for each arm in three scenarios are provided.

1. Scenario 1 (not clinically meaningful for all arms): The true rates of response for T1, T2, T3, T4, and T5 are 40%, 20%, 10%, 10%, and 10%; respectively;
2. Scenario 2 (clinically meaningful for all arms): The true rates of response for T1, T2, T3, T4, and T5 are 60%, 40%, 30%, 30%, and 30%; respectively;
3. Scenario 3 (clinically meaningful for T1 and T5 but not for others): The true rates of response for T1, T2, T3, T4, and T5 are 60%, 20%, 10%, 10% and 30%; respectively;

In all three cases it is assumed that accrual rates are in proportion 2:1:2:2:2 for the 5 arms (T1, T2, T3, T4, and T5). The type I error (false positive rate) of the adaptive design can be viewed as the probability estimate of a positive conclusion at final analysis in Scenario 1 (Table 3-1). The power of the adaptive design is defined as the probability estimate of positive conclusion at final analysis in Scenario 2 (Table 3-2).

The simulation results for Scenario 1, when the effect of Ceritinib is not-clinically meaningful (true response rates equal 40% and 20% and 10% for T1 and T2 and for other 3 arms respectively), show that the probability of stopping at interim for futility varies from 41.7% to 59.9% for different arm. Probability of concluding positive trial at final analysis (i.e., crossed interim and maximum sample size reached) are 11.5%, 13.8%, 12.5%, 11.0%, and 11.0% for T1, T2, T3, T4, and T5 respectively. The overall sample size is approximately 67.

Table 3-1 Simulation results Scenario 1 (ceritinib inactive for all arms)

Arm	True CBR (%)	Probability of stopping at IA for futility (%)	Probability for positive conclusion in final analysis (%)	Average Sample Size
T1	40.0	59.3	11.5	14.1
T2	20.0	41.7	13.8	10.8
T3	10.0	59.9	12.5	14.0
T4	10.0	58.8	11.0	14.1
T5	10.0	59.7	11.0	14.0
Overall				67.0

The simulation-based results for Scenario 2 (Table 3-2), when the truth is that the effect of Ceritinib is clinically meaningful in all arms, illustrate that approximately 69% to 87% of the simulated trial outcomes are positive at final analysis (passes futility interim and reaches the maximum enrollment) for different arm. In addition, the average overall sample size is 91 subjects.

Table 3-2 Simulation results Scenario 2 (ceritinib active for allarms)

Arm	True CBR (%)	Probability of a stopping at IA for futility (%)	Probability for positive conclusion in final analysis (%)	Average Sample Size
T1	60.0	14.2	69.3	18.6
T2	40.0	8.6	77.8	14.1
T3	30.0	5.0	86.8	19.5
T4	30.0	4.6	86.4	19.5
T5	30.0	4.9	86.0	19.5
Overall				91.2

In order to illustrate the design behavior one additional scenario (Scenario 3) is explored. In Scenario 3 the effect of Ceritinib is clinically meaningful for arms T1, T5 and not-clinically meaningful for all other arms (Table 3-3). Table 3-3 shows reasonable operating characteristics (futility at interim and positive results at final analysis) under Scenario 3. The average overall sample size under this scenario is 78.

Table 3-3 Simulation results Scenario 3 (certinib active for T1 and T5 arms)

Arm	True CBR (%)	Probability of stopping at IA for futility (%)	Probability for positive conclusion in final analysis (%)	Average Sample Size
T1	60.0	15.1	71.3	18.5
T2	20.0	35.0	14.9	11.7
T3	10.0	52.2	10.8	14.8
T4	10.0	52.7	11.6	14.7
T5	30.0	13.4	81.0	18.7
Overall				78.4

4 Change to protocol specified analyses

Concomitant medications and significant non-drug therapies prior to and after the start of the study treatment for all patients by arm will not be summarized. For laboratory tests where CTCAE grades are not defined, shift tables using the low/normal/high/(low and high) will not be generated. Only listings will be provided.

5 Appendix

The sections below contain additional details on statistical methodology that will be included in Appendix 16.1.9 (Documentation of Statistical Methods) of the CSR as well as details on programming rules that will be followed to implement the analyses described in Section 2.

5.1 Efficacy endpoints

For further details on efficacy endpoints, see Section 14 (Appendix II) of the protocol. For the evaluation of tumor-response related endpoints, response is assessed by investigator according

to RECIST 1.1 criteria for solid tumors, RECIST 1.1 (for primary endpoint) criteria and RANO [redacted] for GBM or as per Cheson's guidelines for hematological malignancies).

Response and progression evaluation will be performed according to the Novartis RECIST 1.1 guidelines, included in Section 14 (Appendix II) of the LDK378A2407 protocol.

The text below gives more detailed instructions and rules needed for programming of the analyses described in Section 2.

5.1.1 Implementation of RECIST guidelines

Disease progression

PD should only be assigned if it is confirmed by an objective assessment method as per RECIST 1.1 (e.g. radiologic scan, histology for bronchoscopy, photos for skin lesions). If a new lesion is detected using an objective assessment method other than radiologic scan, it should be entered on the 'New lesion' RECIST eCRF with appropriate method (or method='Other').

In particular, discontinuation due to disease progression or death due to progressive disease, without supporting objective evidence (as defined above), will not be considered as PD in the determination of BOR, the derivation of any efficacy endpoint or efficacy analysis.

Change in imaging modality

Per RECIST 1.1, a change in methodology can be defined as either a change in contrast use (e.g. keeping the same technique, like CT, but switching from with to without contrast use or vice-versa, regardless of the justification for the change) or a change in technique (e.g. from CT to MRI, or vice-versa), or a change in any other imaging modality. A change in methodology will result by default in a UNK (unknown) overall lesion response assessment. However, a response assessment other than the Novartis calculated UNK response may be accepted from the investigator if a definitive response assessment can be justified based on the available information. Potential discrepancies between the modality used and overall lesion response reported by the investigator (e.g. change in modality but investigator assessment of response is different from UNK) will be queried during the data validation process.

Determination of missing adequate tumor assessments

For the computation of ORR, patients without any radiological assessment after the start date of study drug will be counted as failure.

Partial or complete responses reported prior to any additional anticancer therapy will be considered for ORR computation irrespective of the number of missed assessments before response. In this section, the 'missing adequate assessment' is defined as assessment not done or assessment with overall lesion response equal to UNK. For the sake of simplicity, the 'missing adequate assessment' will also be referred as 'missing assessment'.

As detailed in Section 14 (Appendix II) of the LDK378A2407 protocol, the PFS censoring and event date options depend on the presence and the number of missing tumor assessments. *For example, an event occurring after two or more missing assessments is censored in the analysis of PFS at the last adequate tumor assessment before the event date.*

An exact rule to determine whether there is none, one or two missing assessments is therefore needed. This rule will be based on the distance between the last adequate tumor assessment date and the event date.

If the distance is larger than threshold D_1 or D_2 then the analysis will assume one or two missing assessments, respectively. The threshold D_1 will be defined as the protocol specified interval between the tumor assessments plus the protocol allowed window around the assessments. Similarly, the threshold D_2 is defined as two times the protocol specified interval between the tumor assessments plus the protocol allowed window around the assessments. In this study, the protocol defined schedule of tumor assessment is every 8 weeks and each assessment is expected to be performed at the scheduled time point plus or minus 1 week, i.e. the window is 2 weeks, then any distance larger than $D_1 = 8+2 = 10$ weeks means one missing assessment and any distance larger than $D_2 = (2*8) + 2 = 18$ weeks means two missing assessments.

The same definition of D_2 will be used to determine the PFS censoring reason. If the distance between the last adequate tumor assessment date and one of the following dates is $\leq D_2$:

1. Analysis cut-off date
2. Start date of further anti-neoplastic therapy
3. Visit date of study treatment discontinuation due to consent withdrawal
4. Visit date of study treatment discontinuation due to loss to follow-up

then the associated censoring reason will be

1. Ongoing
2. New cancer therapy
3. Withdrew consent
4. Loss to follow-up

However, if this distance is larger than D_2, then the censoring reason will be 'Adequate assessment no longer available'.

Non-measurable disease at baseline

As specified in Section 14 (Appendix II) of the LDK378A2407 protocol, the RECIST 1.1 criteria imply that only patients with measurable disease at baseline should be included in the study. If a patient without measurable disease is enrolled, the intent-to-treat (ITT) principle requires including these patients in the analyses. Hence, analyses will be based on FAS including patients with either measurable or non-measurable disease. Therefore, a rule needs to be specified on how to handle these cases.

As specified in Tables 14-7, 14.8, and 14.9 of Section 14 (Appendix II) of the LDK378A2407 protocol, overall lesion response can be derived for patients without measurable disease at baseline as follows (Table 5-1).

Table 5-1 Overall lesion response at each assessment: patients with non-target disease only

Non-target lesions	New Lesions	Overall lesion response
CR	No	CR
Non-CR/Non-PD[1]	No	Non-CR/non-PD
UNK	No	UNK
PD	Yes or No	PD
Any	Yes	PD

[1] In general, the **non-CR/non-PD response** for these patients is considered equivalent to an SD response in endpoint determination.

Missing baseline tumor assessment

As specified in Section 14 (Appendix II) of the protocol, since the timing of PD cannot be determined for patients with missing baseline tumor assessment, these patients are censored in the PFS analysis at the start date of treatment. This rule, however, only applies to the 'PD component' of the PFS assessment.

Patients without baseline tumor assessment who die within D_2 distance from start date of treatment will be counted as having an event in the primary analysis of PFS.

Construction of waterfall graphs

The waterfall graphs will be used to depict the anti-tumor activity. These plots will display the best percentage change from baseline in the sum of the measured diameter of all target lesions for each patient. The proportions of patients with various degrees of tumor shrinkage or growth can then represent a useful efficacy metric.

However, caution needs to be paid to the assessments, where an occurrence of a new lesion or worsening in non-target lesions (resulting in PD as an overall lesion response at given assessment) contradicts the measurements obtained on target lesions. These assessments will not be displayed as bars in the graph. If such a "contradicting" assessment represents the only post-baseline assessment for a patient, then the patient will be represented by a special symbol (e.g. *) in the waterfall graph.

The assessments with unknown target response and also assessments with unknown overall response will be excluded. Patients without any valid assessments will be completely excluded from the graphs.

The total number of patients displayed in the graph needs to be shown and this number will be used as a denominator when calculating the percentages of patients with tumor shrinkage and tumor growth. Footnote will explain the reason for excluding some patients (due to absence of any valid assessment).

All possible assessment scenarios are described in Table 5-2.

Table 5-2 Inclusion/exclusion of assessments used in waterfall graph

Criteria for inclusion/exclusion			Possible source of contradictions	
Target response	Overall lesion response	Include in waterfall	Non-target response	New lesion?
CR/PR/SD	PD	Yes but as * only	PD	any
CR/PR/SD	PD	Yes but as * only	any	Yes
UNK	UNK or PD	No	any	any
CR/PR/SD	UNK	No	UNK	No
CR/PR/SD	CR/PR/SD	Yes as a bar	SD/IR	No
PD	PD	Yes as a bar	any	any

The following algorithm will be used to construct the graph:

1. Select "valid" post-baseline assessments to be included, i.e. for each patient and each assessment repeat the following four steps:

1.1. Check the target lesion response and overall lesion response at each assessment. If at least one of them is UNK then exclude the whole assessment. Otherwise, go to step 1.2.

1.2. Check the overall lesion response. If PD, then go to step 1.3. Otherwise, go to step 1.4

1.3. Check target response. If PD, then go to step 1.4. Otherwise flag the assessment ✶.

1.4. Calculate the % change from baseline in target lesions.

2. For each patient, go through all valid assessments identified in step 1 and find the assessment with best % change from baseline in target lesions. The "best" means best for the patient, i.e. the largest shrinkage or if a patient only has assessments with tumor growth take the assessment where the growth is minimal. (*Example 1*: Patient 1 has the following % changes from baseline at assessments 1, 2, 3, 4 and 5, respectively: -10%; -25%; -13%; -4% and +6%. His/her best % change is then -25%. *Example 2*: Patient 2 has the following % changes from baseline at assessments 1, 2 and 3, respectively: +5%; +18% and +35%. His/her best % change is then +5%.
3. Construct the waterfall graph displaying the best % change from baseline for each patient. Patients having only ✶ flagged assessment(s) will be displayed separately.

Investigator assessment will be used in the construction of waterfall plot.

The recommended way of the display from left to right is:

1. Bars under the horizontal axis representing tumor shrinkage
2. Bars above the horizontal axis representing tumor growth
3. "Zero" bars with ✶ symbol representing patients with contradiction

5.1.2 Response assessment in neuro-oncology (RANO) criteria for GBM

Antitumor response will be primarily evaluated by the Response Assessment in Neuro-Oncology (RANO) working group (Wen et al 2010) criteria in this study. The RANO Criteria updates its established predecessor, the modified Macdonald Criteria (Macdonald et al 1990), by adding assessment of non-enhancing lesions.

Patients will undergo Contrast MRI assessments for response evaluation starting at Week 8 and every 8 weeks thereafter to evaluate brain lesions, as outlined in the Visit schedule [Table 7-1] and [Table 7-2] in the protocol.

The following components will be taken into account when assessing a patient's overall response at an individual evaluation.

- Tumor evaluation eCRF page for measureable enhancing lesions (T1-Gd+)
- Tumor evaluation eCRF page for non-measurable enhancing lesions (T1-Gd+)
- Tumor evaluation eCRF page for non-enhancing lesions (T2/FLAIR)
- Tumor evaluation eCRF page for new lesion
- Concomitant medication eCRF page for steroid usage
- Clinical status eCRF page for ECOG performance status (WHO) and other clinical evaluation finding as per investigator
- Overall response eCRF page for response category (CR/PR/PD/SD/NA)

5.1.2.1 Antitumor effect - definitions

Evaluable for toxicity

All participants who receive at least one dose of study treatment will be evaluable for toxicity from the time of their first treatment.

Evaluable for objective response

Only those participants who have measurable disease present at baseline (cycle 1, day 1 scan) and have received at least one dose of therapy will be considered evaluable for response. These participants will have their response classified according to the definitions stated below. (Note: Participants who exhibit objective disease progression or die prior to the end of cycle 1 will also be considered evaluable.)

Measurable disease

Bidimensionally, contrast-enhancing, measurable lesions with clearly defined margins by CT or MRI scan, with a minimal diameter of 10mm, and visible on 2 axial slices which are at least 5 mm apart with 0 mm skip. Measurement of tumor around a cyst or surgical cavity, if necessary, requires a minimum thickness of 3 mm. If there are multiple measurable lesions, the investigator must choose a minimum of two and a maximum of five of the largest lesions to be followed before a participant is entered on study. The remaining lesions will be considered non-measureable for the purpose of objective response determination. Unless progression is observed, objective response can only be determined when all measurable and non-measurable lesions are assessed.

Non-measurable evaluable disease

Unidimensionally measurable lesions, masses with margins not clearly defined, lesions with maximal diameter < 10mm.

5.1.2.2 Response/progression categories

Complete response (CR)

All of the following criteria must be met:

a. Complete disappearance of all enhancing measurable and non-measurable disease sustained for at least 4 weeks. In the absence of a confirming scan 4 weeks later, this scan will be considered only stable disease.
b. No new lesions.
c. All measurable and non-measurable lesions must be assessed using the same techniques as baseline.
d. Participants must be on no steroids or on physiologic replacement doses only.
e. Stable or improved non-enhancing (T2/FLAIR) lesions
f. Stable or improved clinically, for clinical signs and symptoms present at baseline and recorded to be disease related

Participants with non-measurable disease cannot have a complete response. The best response possible is stable disease.

Partial response (PR)

All of the following criteria must be met:

a. Greater than or equal to 50% decrease compared to baseline in the sum of products of perpendicular diameters of all measurable enhancing lesions sustained for at least 4 weeks. In the absence of a confirming scan 4 weeks later, this scan will be considered only stable disease.
b. No progression of non-measurable disease.
c. No new lesions.
d. All measurable and non-measurable lesions must be assessed using the same techniques as baseline.
e. The steroid dose at the time of the scan evaluation should be no greater than the dose at time of baseline scan.
f. Stable or improved non-enhancing (T2/FLAIR) lesions on same or lower dose of corticosteroids compared to baseline scan.
g. Stable or improved, for clinical signs and symptoms present at baseline and recorded to be disease related clinically.

Participants with non-measurable disease cannot have a partial response. The best response possible is stable disease.

Progressive disease (PD)

The following criterion must be met:

a. 25% increase in sum of the products of perpendicular diameters of enhancing lesions (over best response or baseline if no decrease) on stable or increasing doses of corticosteroids **and/or one or more of the of the following:**
b. Significant increase in T2/FLAIR non-enhancing lesion on stable or increasing doses of corticosteroids steroids compared to baseline scan or best response following initiation of therapy, not due to co-morbid events (radiation therapy, demyelination, ischemic injury, infection, seizures, post-operative changes, or other treatment effects).
c. Any new lesion
d. Clear clinical deterioration not attributable to other causes apart from the tumor (e.g. seizures, medication side effects, complications of therapy, cerebrovascular events, infection, etc.). The definition of clinical deterioration is left to the discretion of the investigator but it is recommended that a decline in the ECOG performance score (WHO) from 0 or 1 to 2 or 2 to 3 would be considered neurologic deterioration, unless attributable to co-morbid events or changes in corticosteroid dose.
e. Failure to return for evaluation due to death or deteriorating condition.
f. Clear progression of non-measurable disease

Stable disease (SD)

All of the following criteria must be met:

a. Does not qualify for CR, PR, or progression.
b. All measurable and non-measurable sites must be assessed using the same techniques as baseline.
c. Stable non-enhancing (T2/FLAIR) lesions on same or lower dose of corticosteroids compared to baseline scan. In the event that the corticosteroid dose has been increased, the last scan considered to show stable disease will be the scan obtained when the corticosteroid dose was equivalent to the baseline dose.
d. Stable clinically.

Unknown response status

Progressive disease has not been documented and one or more measurable or non-measurable lesions have not been assessed.

These RANO Response Criteria are also summarized in Table 5-3:

Table 5-3 Summary of the RANO response criteria

	CR	PR	SD	PD#
T1-Gd +	None	≥50% decrease	<50% decrease but <25% increase	≥25% increase*
T2/FLAIR	Stable or decrease	Stable or decrease	Stable or decrease	Increase*
New Lesion	None	None	None	Present*
Corticosteroids	None	Stable or decrease	Stable or decrease	NA**
Clinical Status	Stable or improve	Stable or improve	Stable or improve	Deterioration*
Requirement for Response	All	All	All	Any*

CR=complete response; PR=partial response; SD=stable disease; PD=progressive disease
*: Progression when this criterion is met **: Increase in corticosteroids alone will not be taken into account in determining progression in the absence of persistent clinical deterioration

5.1.2.3 Methods for evaluation of measurable disease

All measurements should be taken and recorded in metric notation, using a ruler, calipers, or digital measurement tool. All baseline evaluations should be performed as closely as possible to the beginning of treatment and never more than 14 days before the beginning of the treatment.

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up.

5.1.2.4 Evaluation of best response

The best overall response is the best response recorded from the start of the treatment until disease progression (taking as reference for progressive disease the smallest measurements recorded since the treatment started). If a response recorded at one scheduled MRI does not persist at the next regular scheduled MRI, the response will still be recorded based on the prior scan, but will be designated as a non-sustained response. If the response is sustained, i.e. still present on the subsequent MRI, it will be recorded as a sustained response, lasting until the time of tumor progression. Participants without measurable disease may only achieve SD or PD as their best "response."

5.1.2.5 Neurological exam

Although not used for determining response, it is useful to evaluate changes in the neurological exam compared to the previous exam. The following scale may be used:

+2	Definitely better
+1	Possibly better
0	Unchanged
-1	Possibly worse
-2	Definitely worse

5.1.2.6 Performance status

Participants will be graded according to WHO performance status.

5.1.2.7 Overall survival time

From date of first dose (date of first post-surgery treatment for participants in Dose Level 1) to date of death due to any cause.

5.1.2.8 Progression-free survival time:

From date of first dose (date of first post-surgery treatment for participants in Dose Level 1) to date of progression or death. Participants who stop treatment for causes other than progression may be censored if other therapy is initiated or if regular assessments for assessing progression are no longer available.

5.1.3 Guidelines for efficacy evaluation in lymphoma studies (based on Cheson response criteria). International Working Group guidelines for hematological malignancies

Disease assessments will be based on the International Working Group response criteria (Cheson 1999), and the International Harmonization Project revised response criteria (Cheson et al 2007b). Further clarification on these criteria has been published by (Cheson 2007a).

5.1.3.1 Definitions and criteria for normalization

A lesion is categorized based on the location as:

- **Nodal lesion,**
- **Extranodal lesion**, if it is located in organs other than lymph node or nodal mass, but including spleen and liver.

5.1.3.2 Measurability of tumor lesions at baseline

All tumor lesions/lymph nodes will be categorized as measurable or non-measurable as follows:

5.1.3.2.1 Measurable nodal and extranodal lesions

A lesion will be called **measurable** if it can be measured accurately in 2 perpendicular dimensions and:

- For nodal lesion, if the long axis is > 15 mm, regardless of the length of the short axis,
- For extranodal lesion, if the long and short axes are ≥ 10 mm.

Patients should have at least one measurable nodal lesion greater than 20 mm in the long axis. In cases where the patient has no measurable nodal lesions greater than 20 mm in the long axis at Screening, then the patient must have at least one measurable extranodal lesion.

5.1.3.2.2 Classification of lymph nodes

Lymph nodes are classified according to their size and/or relationship to the disease:

- A lymph node meeting the measurability requirement above will constitute a **measurable nodal lesion**.
- A lymph node not meeting the measurability requirement but with long axis > 15 mm (e.g. short axis cannot be measured accurately) will constitute a **non-measurable nodal lesion**.

- A lymph node not meeting the measurability criteria but with a size of 11 mm to 15 mm in the long axis and > 10 mm in the short axis will be checked for relationship to disease:
  - If it is thought to be disease related, it will constitute a **non-measurable nodal lesion**.
  - If it is not thought to be disease related, it will constitute an **abnormal lymph node** but not a lesion.
- All other lymph nodes will be considered normal and will not constitute nodal lesions.

5.1.3.2.3 Criteria for normalization of lesions

The normalization of lesions is defined as follows:

- A measurable nodal lesion must become ≤ 15 mm in long axis to be considered normalized.
- A non-measurable nodal lesion must decrease to ≤ 10 mm in the short axis and be ≤ 15 mm in long axis to be considered normalized.
- An extranodal lesion must disappear completely (assigned a size of 0 mm x 0 mm) to be considered normalized.

5.1.3.3 Specification by methods of measurement

5.1.3.3.1 Measurement of lesions

All radiological measurements should be taken in two perpendicular dimensions and recorded in metric notation, using a ruler or calipers.

PET

Visual assessment currently is considered adequate for determining whether a PET scan is positive, and use of the standardized uptake value is not necessary.[1] In brief, a positive scan is defined as focal or diffuse FDG uptake above background in a location incompatible with normal anatomy or physiology, without a specific standardized uptake value cutoff.[1] Other causes of false-positive scans should be ruled out. Exceptions include mild and diffusely increased FDG uptake at the site of moderate- or large-sized masses with an intensity that is lower than or equal to the mediastinal blood pool, hepatic or splenic nodules 1.5 cm with FDG uptake lower than the surrounding liver/spleen.

CT scan (or MRI)

For optimal evaluation of patients, the same methods of assessment and technique should be used to characterize each identified and reported lesion at Screening and during follow-up. Contrast-enhanced CT of chest, abdomen and pelvis should preferably be performed using a 5mm slice thickness with a contiguous reconstruction algorithm. CT/MRI scan slice thickness should not exceed 8 mm cuts using a contiguous reconstruction algorithm. If at Screening a patient is known to be allergic to CT contrast or develops allergy during the trial, the following change in imaging modality will be accepted for follow up: a non-contrast CT of chest (MRI not recommended due to respiratory artifacts) plus contrast-enhanced MRI of abdomen and pelvis.

A change in methodology can be defined as either a change in contrast use (e.g. keeping the same technique, like CT, but switching from with to without contrast use or vice-versa, regardless of the justification for the change) or a change in technique (e.g. from CT to MRI, or vice-versa), or a change in any other imaging modality. A change in methodology will result by default in an "Unknown" overall radiological response assessment. However, another overall radiological response than the Novartis calculated "Unknown" response may be accepted from the investigator if a definitive overall radiological response can be justified to be based on the available information.

In order to calculate the sum of the product of the diameters (SPD) of all index lesions (or extranodal lesions), their size must be entered throughout the study.

Actual lesion measurements should be entered on the corresponding eCRFs. If, during the course of the study, either of the perpendicular diameters of a lesion cannot be reliably measured because of its small size, it is recommended to enter the minimum limit of detection as the diameter size (e.g. 5 mm for spiral CT). In other cases when, during the course of the study, the diameter cannot be reliably measured for reasons other than its size (i.e. borders of the lesion are confounded by neighboring anatomical structures), no measurement should be entered and the lesion cannot be evaluated.

If lesions become confluent over time, it is recommended to measure them as one lesion, report the overall diameters to one of the lesions and assign 0 mm x 0 mm to each of the other previously measured lesions. If a lesion splits during the study, each sub-lesion should be measured separately for all subsequent assessments and all sub-lesions contribute to the SPD.

Bone marrow assessment

Documentation of status of bone marrow involvement by lymphoma based on prior bone marrow biopsy or aspirate findings is required at Screening for all patients.

If no such documentation is available then a bone marrow biopsy or aspirate should be performed at Screening.

If bone marrow involvement is assessed by biopsy, the biopsy sample should have a goal of > 20 mm unilateral core. If the biopsy sample is indeterminate by morphology (immunohistochemistry), then flow cytometry may be performed on bone marrow aspirate to confirm the findings.

Physical examination and assessment of B-symptoms

Skin lesions, if the size is ≥ 20 mm in at least one diameter, must be histologically confirmed for lymphoma involvement (the investigational site must document the histological confirmation (yes or no) on the corresponding eCRF) and photographed including a ruler (color photography using digital camera). Tumor assessment will be performed and results will be recorded on the corresponding eCRF at Screening and at Day 1 of every cycle (±4 days) after first dose of study drug.

B-symptoms are of importance in determining prognosis and should resolve completely in patients who have achieved complete response. B-symptoms in lymphoma patients are disease related clinical symptoms and are not caused by anticancer therapy (or drug toxicity).

B-symptoms are defined as follows:

- Significant unexplained fever (≥ 38°C),
- Unexplained, recurrent drenching night sweats
- Unexplained loss of > 10% body weight within the previous 6 months, as assessed and reported (present vs. absent) by the Investigator.

5.1.3.4 Evaluation of radiological response

For the sake of simplicity, complete remission and complete response will both be referred to as complete response.

Definitions of Response for Lymphoma patients are listed in Table 5-5. To evaluate disease response to treatment, all index and non-index lesions will be followed and assessed throughout the study. At each assessment, response is evaluated separately for the **index lesions** (Table 5-8) and **non-index lesions** (Table 5-7) identified at Screening, then a combined overall radiological response is determined (Table 5-10).

Table 5-4 Response definition for lymphoma

Response	Definition	Nodal Masses	Spleen. Liver	Bone Marrow
CR	Disappearance of all evidence of disease	a. FDG-avid or PET positive prior to therapy; mass of any size permitted if PET negative b. Variably FDG-avid or PET negative; regression to normal size on CT	Not palpable, nodules disappeared	Infiltrate cleared on repeat biopsy; if indeterminate by morphology, immunohistochem is try should be negative
PR	Regression of measurable disease and no new sites	≥ 50% decrease in SPD of up to 6 largest dominant masses; no increase in size of other nodes a. FDG-avid or PET positive prior to therapy; one or more PET positive at previously involved site b. Variably FDG-avid or PET negative; regression on CT	≥ 50% decrease in SPD of nodules (for single nodule in greatest transverse diameter); no increase in size of liver or spleen	Irrelevant if positive prior to therapy; cell type should be specified
SD	Failure to attain CR/PR or PD	a. FDG-avid or PET positive prior to therapy; PET positive at prior sites of disease and no new sites on CT or PET b. Variably FDG-avid or PET negative; no change in size of previous lesions on CT		
Relapsed disease or PD	Any new lesion or increase by ≥ 50% of previously involved sites from nadir	Appearance of a new lesion(s) > 1.5 cm in any axis, ≥50% increase in SPD of more than one node, or ≥ 50% increase in longest diameter of a previously identified node > 1 cm in short axis Lesions PET positive if FDG-avid lymphoma or PET positive prior to therapy	> 50% increase from nadir in the SPD of any previous lesions	New or recurrent involvement

5.1.3.5 Evaluation of index lesions (nodal and extranodal)

5.1.3.5.1 When index nodal lesions are not in complete response

The response for index lesions is evaluated by calculating the Sum of the Products of Diameters (SPD) of all index lesions (see Table 5-6), except when there is a Complete Response for index nodal lesions (i.e. complete normalization of all index nodal lesions) (see Section 5.1.3.2.3).

Table 5-5 Radiological status based on SPD calculation for all index lesions

Response Criteria[1]	Evaluation of index lesions
Complete Response (CR)	See Table 5-8 below (not based on SPD calculation for all index lesions)
Partial Response (PR)	At least 50% decrease from Screening in the SPD of all index lesions
Stable Disease (SD)	Failure to attain the criteria needed for CR or PR and failure to fulfill the criteria for PD
Progressive Disease (PD)	At least a 50% increase from nadir2 in the SPD of all index lesions

[1] At each assessment (if the index nodal lesions are not in CR status), the response status based on SPD calculation will be first assessed for meeting PD status criteria, then PR status and SD status.
[2] Nadir is defined as the smallest sum of the product of the diameters of all index lesions recorded so far, at or after Screening.

5.1.3.5.2 When index nodal lesions are in complete response

When there is a Complete Response for index nodal lesions (i.e. complete normalization of all index nodal lesions as defined in Section 5.1.3.2.3: all index lesion ≤ 15 mm in long axis), the SPD for these index nodal lesions may not be equal to zero and therefore a calculation of a SPD for all index lesions may be misleading. Therefore, by default, a specific response for extranodal index lesions needs to be evaluated, based on the SPD calculation restricted to all index extranodal lesions only (see Table 5-7).

Table 5-6 Radiological response criteria for index extranodal lesions in case of CR in index nodal lesions

Response Criteria[1]	Evaluation of index extranodal lesions
Complete Response (CR)	Complete disappearance of all index extranodal lesions
Partial Response (PR)	At least 50% decrease from Screening in the SPD restricted to all index extranodal lesions
Stable Disease (SD)	Failure to attain the criteria needed for CR or PR and failure to fulfill the criteria for PD
Progressive Disease (PD)	At least a 50% increase from nadir[2] in the SPD restricted to all index extranodal lesions

[1] At each assessment, response will be first assessed for meeting CR status. If CR status is not met, response will be assessed for PD status, then PR status and SD status.
[2] Nadir is defined as the smallest sum of the product of the diameters restricted to all index extranodal lesions recorded so far, at or after Screening.

The algorithm for evaluating the response integrating index extranodal lesions and the SPD calculated on all index lesions (where appropriate) provides an overall response for index lesions.

5.1.3.5.3 Evaluation of response for all index lesions

The evaluation of response for all index lesions is based on the combination of the response for index nodal lesions (CR or non-CR), the response for index extranodal and the status based on the SPD calculated on all index lesions (nodal and extranodal), as described in Table 5-8.

Table 5-7 Radiological response for index lesions

Response for index nodal lesions[1]	Response for index extranodal lesions[1]	Status based on SPD calculation for all index lesions	Response for index lesions
CR	CR	Not calculated	CR
CR	SD/ PR	Not calculated	PR
CR	PD	PD	PD
CR	PD	PR	PR
CR	PD	SD	SD
Non-CR	Not evaluated	PD	PD
Non-CR	Not evaluated	PR	PR
Non-CR	Not evaluated	SD	SD

[1] If no index nodal lesions are present at Screening, then index lesions response is equal to the index extranodal lesions response. A similar rule applied if no index extranodal lesions are present at Screening, then index lesions response is equal to the index nodal lesions response.

In case of missing measurements of any of the index lesions, the radiological response for index lesions at that assessment will be "Unknown (UNK)", unless progression was seen.

All lesions must have been measured with the same method as the one used at Screening, otherwise the radiological response for index lesions at that assessment will be "Unknown (UNK)".

5.1.3.5.4 Evaluation of non-index lesions (including nodal, splenic and/or hepatic nodules and other extranodal lesions)

At each reassessment, a non-index lesion (or a group of non-index lesions) will be given one of the following designations:

- Normalization (non-index nodal lesion has regressed to normal size; non-index extranodal lesion is no longer present). Normalization of non-index nodal lesions should be determined based on their size at Screening.
- Improved, stable or worsened, but without unequivocal evidence of disease progression (non-index lesion is present but there is not sufficient worsening to declare PD based on the existing non-index lesions).
- Unequivocal evidence of disease progression (worsening of existing non-index lesions is sufficient to declare PD).
- Not assessed.

Then, this status for each non-index lesion (or group of non-index lesions) will lead to a global response for non-index lesions (Table 5-9):

Table 5-8 Response criteria for non-index lesions (nodal, splenic and/or hepatic nodules and other extranodal lesions)

Response Criteria	Evaluation of non-index lesions
Complete Response (CR)	Complete normalization of all non-index nodal and extranodal lesions: Radiological regression to normal size of all lymph nodes and complete disappearance of all extranodal (including splenic and/or hepatic nodules) lesions
Stable Disease (SD)	Failure to attain the criteria needed for CR and failure to fulfill the criteria for PD
Progressive Disease (PD)	Unequivocal disease progression of any existing non-index lesions (nodal or extranodal)

In case of a missing status of any of the non-index lesions, the radiological response for non-index lesions at that assessment will be "Unknown (UNK)", unless progression was seen.

All lesions must have been measured with the same method as the one used at Screening, otherwise the radiological response for non-index lesions at that assessment will be "Unknown (UNK)".

5.1.3.6 New lesions

The appearance of

- any new nodal lesion >15 mm in any axis. New nodal lesion is defined by:
  - either a previously normal lymph node becoming > 15 mm in any axis,
  - or a previously identified abnormal lymph node showing an increase of at least 50% in the long axis,
  - as assessed by investigator

OR

- any discrete extranodal (including splenic and/or hepatic nodules) lesions reliably appearing on CT scan or MRI after Screening.

is always considered as Progressive Disease (PD) and has to be recorded as a new lesion in the appropriate module of the eCRF. Determination of new lymphoma involvement in organs other than lymph nodes or liver or spleen should be confirmed histologically and the site must document that in a comment to the corresponding eCRF.

5.1.3.6.1 Overall radiological response

Overall radiological response is calculated as shown in Table 5-10.

Table 5-9 Overall radiological response at each assessment

Index lesions	Non-index lesions[1]	New lesions	Overall radiological response
CR	CR	No	CR
CR	SD	No	PR
PR	CR or SD	No	PR
SD	CR or SD	No	SD
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

[1] If no non-index lesions are present at Screening, then this column is not used in evaluating overall radiological response.

If the evaluation of any of the index or non-index lesions identified at Screening could not be made during follow-up or if the index or non-index response is "Unknown (UNK)", the overall response status at that assessment must be "Unknown (UNK)" unless progression or a new lesion was seen.

5.1.3.6.2 Evaluation of overall disease response

The evaluation of overall disease response at each assessment is a composite of the individual radiological responses (index and non-index lesions, new lesions), laboratory test (bone marrow) and clinical responses (lymphoma related clinical symptoms).

5.1.3.6.3 Bone marrow re-assessment at time of radiological CR

In order to confirm a Complete disease response (CR), bone marrow biopsy or aspirate may be required when a radiological CR has been achieved. Details are provided in the Study Protocol. The infiltrate of lymphoma in bone marrow must have cleared on repeat bone marrow biopsy or aspirate. Patients who achieve a CR by other criteria but who have persistent morphologic positive or inconclusive bone marrow involvement will be considered partial responders. New or recurrent bone marrow involvement anytime during the follow up will be considered PD. Bone marrow biopsy or aspirate will be performed after the first assessment of CR or when clinically indicated.

The biopsy sample of bone marrow must be adequate (with a goal of > 20 mm unilateral core). If the sample is indeterminate by morphology, it should be negative by immunohistochemistry.

5.1.3.6.4 Overall disease response

If a patient has an overall radiological response of CR then this response must be confirmed by bone marrow biopsy or aspirate (if required as per the Study Protocol), presence of normal liver and spleen size, and evaluation of lymphoma related B-symptoms. The patient's overall response will be calculated as follows:

A patient will be deemed to have overall disease response of CR if bone marrow biopsy or aspirate becomes negative for tumor involvement (if the bone marrow was involved by lymphoma at Screening) and the liver and spleen are normal in size and there are no lymphoma related B-symptoms in addition to radiological CR.

If assessments of any of the following: lymphomatous infiltration of bone marrow (If required as per the Study Protocol), or evaluation of B-symptoms is not done, unknown or indeterminate or B-symptoms are still present when the overall radiological response is assessed as CR or the liver or spleen are enlarged, then the overall disease response will be assessed as PR until evaluation of these factors have shown normalized results and recorded on the corresponding eCRF.

For patients whose radiological response is anything other than CR, assessment of bone marrow, liver, spleen and B-symptoms will not be required in evaluating overall response and overall disease response is the same as radiological response. However any new or recurrent bone marrow involvement at any time during follow-up will be considered PD.

Of note, appearance of B-symptoms or enlarged spleen or liver will not in themselves constitute documentation of progression. They are however expected to be associated with progressive disease. Every effort should be made to document that evidence radiologically and report the corresponding tumor assessments. Such tumor assessments are expected to be performed within 2 months of appearance of B-symptoms or enlarged spleen or liver.

5.1.3.7 References (available upon request)

Cheson BD (2007a) The international harmonization project for response criteria in lymphoma clinical trials. Hematol Oncol Clin N Am 21:841-854.

Cheson BD (2009) The case against heavy PETing. J Clin Oncol 27:1742-1743.

Cheson BD, Horning SJ, Coiffier B, et al (1999) Report of an International Workshop to standardize response criteria for non-Hodgkin's lymphomas. J Clin Oncol 17:1244-1253.

Cheson BD, Pfistner B, Juweid ME, et al (2007b) Revised response criteria for malignant lymphoma. J Clin Oncol 25:579-586.

FDA Guideline (2005) Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, April 2005.

5.2 Imputation rules

5.2.1 Study drug

For patients not known to have died prior to the cut-off date:

- All events with start date before or on the cut-off date, and with end date missing or after the cut-off date will be reported as “continuing at the cut-off date”.
- This approach applies, in particular, to AEs and concomitant medication reports. For these events, the end date will not be imputed and therefore will not appear in the listings.

For patients known to have died prior to or on the cut-off date:

- All events with start date before or on the cut-off date, and with end date missing or after the cut-off date will have the end date imputed to the death date.

- This approach applies, in particular, to AEs and concomitant medication reports. For these events, the imputed end date will not appear in the listings.

If imputation of an end date is required for a specific analysis (e.g. for a dose administration record with missing end date or last date of study drug after the cut-off date), the end date will be imputed to the cut-off date for the purpose of calculating duration of exposure to study drug and dose intensity. The imputed date will be displayed and flagged in the listings.

5.2.2 AE date imputation

Date imputation is the creation of a new, complete date from a partial one according to an agreed and acceptable algorithm. Missing date for AE will be handled according to rules specified below. A partial date is simply an incomplete date e.g. DDOCT2001: the days are missing from this DDMMMYYYY date.

Partial AE start dates, if left partial, would ultimately mean the following

It would not be possible to place the AE in time.

Therefore the treatment/dosage at the time of the event would be unknown.

Therefore the event could not be reported/summarized appropriately – if at all.

Therefore it is important to perform date imputation to ensure that as many data events are represented as correctly as possible. Of course partial and/or missing dates should *also* be caught as edit checks and passed back to the investigator for resolution.

There **will be no** attempt to impute the following:

- **Missing** AE start dates
- AE start dates **missing the year**
- Partial/missing AE **end dates**

The following Table 5-11 explains the abbreviations used.

Table 5-10 AE/treatment date abbreviations

	Day	Month	Year
Partial Adverse Event Start Date	<not used>	AEM	AEY
Treatment Start Date (TRTSTD)	<not used>	TRTM	TRTY

The following matrix Table 5-12 describes the possible combinations and their associated imputations. In the light grey boxes the upper text indicates the imputation and the lower text the relationship of the AE start date to the treatment start date (TRTSTD).

Table 5-11 AE partial date imputation algorithm

	AEM MISSING	AEM < TRTM	AEM = TRTM	AEM > TRTM
AEY MISSING	NC Uncertain	NC Uncertain	NC Uncertain	NC Uncertain
AEY < TRTY	(D) Before TRTSTD	(C) Before TRTSTD	(C) Before TRTSTD	(C) Before TRTSTD
AEY = TRTY	(B) Uncertain	(C) Before TRTSTD	(B) Uncertain	(A) After TRTSTD
AEY > TRTY	(E) After TRTSTD	(A) After TRTSTD	(A) After TRTSTD	(A) After TRTSTD

The following Table 5-13 is the legend to the above table.

Table 5-12 AE/treatment date relationship and imputation legend

Relationship	
Before TRTSTD	Indicates AE start date prior to Treatment Start Date
After TRTSTD	Indicates AE start date after Treatment Start Date
Uncertain	Insufficient to determine the relationship of AE start date to Treatment Start Date
Imputation Calculation	
NC / Blank	No convention/imputation
(A)	01MONYYYY
(B)	TRTSTD+1
(C)	15MONYYYY
(D)	01JULYYYY
(E)	01JANYYYY

The following Table 5-14 gives a few examples.

Table 5-13 AE imputation example scenarios

Partial AE start date	Treatment start date	Relationship	Imputation Calculation	Imputed Date
12mmyyyy	20OCT2001	Uncertain	NC	<blank>
ddmmm2000	20OCT2001	Before	(D)	01JUL2000
ddmmm2002	20OCT2001	After	(E)	01JAN2002
ddmmm2001	20OCT2001	Uncertain	(B)	21OCT2001
ddSEP2001	20OCT2001	Before	(C)	15SEP2001

Partial AE start date	Treatment start date	Relationship	Imputation Calculation	Imputed Date
ddOCT2001	20OCT2001	Uncertain	(B)	21OCT2001
ddNOV2001	20OCT2001	After	(A)	01NOV2001

5.2.3 Concomitant medication date imputation

The imputation of the start date of concomitant medication will follow the same conventions as for AE date. Partial concomitant medication end dates will not be imputed.

5.2.4 Incomplete date of initial diagnosis of cancer and date of most recent recurrence

Missing day is defaulted to the 15th of the month and missing month and day is defaulted to 01-Jan.

5.2.5 Incomplete date for anti-neoplastic therapies

5.2.5.1 Prior therapies date imputation

Prior therapies

Start date:

The same rule which is applied to the imputation of AE/concomitant medication start date will be used with the exception that for scenario (B) will be replaced to be 'start date of study drug -1'.

End date:

Imputed date = min (start date of study drug, last day of the month), if day is missing;

Imputed date = min (start date of study drug, 31DEC), if month and day are missing.

If the end date is not missing and the imputed start date is after the end date, use the end date as the imputed start date.

If both the start date and the end date are imputed and if the imputed start date is after the imputed end date, use the imputed end date as the imputation for the start date.

5.2.5.2 Post therapies date imputation

Post therapies

Start date:

Imputed date = max (last date of study drug + 1, first day of the month), if day is missing;

Imputed date = max (last date of study drug + 1, 01JAN), if day and month are missing.

End date: No imputation.

5.2.6 Incomplete assessment dates for tumor assessment

All investigation dates (e.g. X-ray, CT scan) must be completed with day, month and year.

If one or more investigation dates are incomplete but other investigation dates are available, the incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the latest of all investigation dates (e.g. X-ray, CT-scan) if the overall lesion response at that assessment is CR/PR/SD/UNK. Otherwise, if overall lesion response is PD, the assessment date is calculated as the earliest date of all investigation dates at that evaluation number. If all measurement dates have no day recorded, the 1st of the month is used.

If the month is not completed, for any of the investigations, the respective assessment will be considered to be at the date which is exactly between the previous and the following assessment. If both a previous and following assessments are not available, this assessment will not be used for any calculations.

5.2.7 Incomplete date for death

All dates must be completed with day, month and year.

If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) last contact date (excluding the date of death) and the following:

- Missing day: 15th day of the month and year of death
- Missing day and month: July 1st of the year of death

5.2.8 Incomplete dates for disease progression on medication prior to the start of study drug.

If day of progression associated with the 'Prior Antineoplastic Therapy – Medication' page is missing then the imputed PD date is:

= min (midpoint between the end date of the prior medication and the end of the month, start date of LDK), if end date of prior medication is in the same month as the PD date.

= min (15th of the month of the PD date, start date of LDK), if end date of prior medication is in a month prior to the PD date.

= 15th of the month of the PD date, if end date of medication is in a month after the PD date.

If both day and month of progression associated with the 'Prior Antineoplastic Therapy – Medication' page is missing then the imputed PD date is:

= min (midpoint between the end date of the prior medication and the end of the year, start date of LDK) if end date of prior medication is in same year as the PD date

= min (July 1 of the year of the PD date, start date of LDK), if end date of prior medication is in a year prior to the PD date.

= July 1 of the year of the PD date, if end date of prior medication is in a year after the PD date.

Completely missing PD dates will not be imputed. For the midpoint calculation, if odd days in between (e.g last dose of criztonib is 27 June 2012, and end of the month is 30 June 2012), then use the next day from the midpoint calculation (e.g midpoint is 29 June 2013).

5.2.9 Incomplete dates for last dose of study drug

Scenario 1

If the last date of study drug is after the cut-off date or is completely missing and there is no end of treatment eCRF page and no death date the patient should be considered to be on-going and use the cutoff date for the analysis as the last dosing date

Scenario 2

If the last date of study drug is completely or partially missing and there is EITHER an end of treatment eCRF page OR a death date available then imputed last dose date:

= 31DECYYYY, if only Year is available and Year < Year of min (EOT visit date, death date)

= Last day of the month, if both Year and Month are available and Year = Year of min (EOT visit date, death date) and Month < the month of min (EOT visit date, death date)

= min (EOT visit date, death date), for all other cases

The imputed date will be compared with start date of study drug.

If the imputed date < start date of study drug, then last date of study drug is set to start date of study drug;

Otherwise, use the imputed date.

5.3 Safety evaluations

The text below gives more detailed instructions and rules needed for programming of the analyses described in Section 2.

5.3.1 Multiple assessments within post-baseline visits

For all analyses regarding abnormal assessments or analyses based on worst or best post-baseline value (laboratory, ECGs, vital signs, ECOG performance status), all post-baseline values will be included (scheduled, unscheduled, repeat). All unscheduled and repeat measurements will be included in listings.

Laboratory Data

For laboratory data, assessments can be collected from both local and central laboratory on the same date. For shift tables using CTC grades to compare baseline to the worst post-baseline value, the assessment with worst post-baseline value is used for analyses irrespective of the source. For LFT summaries, where concurrent measurements are used in the calculation of number and percentage of patients with worst post-baseline values, the assessment with worst post-baseline value is used (since worst values are based on the largest ratio of lab value to its

ULN for each patient) although the worst values for the different parameters may be coming from different laboratories.

ECGs

For all patients, 3 ECGs are targeted to be measured at the protocol-defined (nominal) time-points. If a patient has more than one measurement at a nominal time point, the average of all available measurements associated with the nominal time point will be used for the analyses.

5.3.2 Baseline

As defined in Section 2.1.1, the last available assessment before or on the date of start of study drug is defined as "baseline" value or "baseline" assessment.

Laboratory data

If both central and local laboratory assessments were performed on the same date and corresponding to the baseline assessment date, then the central laboratory assessment will be used for the calculation of baseline.

ECGs

Baseline for ECG measurement is the average of the pre-dose replicate measurements on the baseline day. Unscheduled assessments will not be included in the calculation of the average. Study day 1 scheduled pre-dose ECGs will be considered to have been obtained prior to study drug administration if dosing time or ECG time is missing.

If a scheduled pre-dose measurement actually occurred post-dose, then the corresponding measurement will be treated and analyzed similar to an unscheduled post-dose measurement. For unscheduled assessments on study day 1,

- if dosing time is non-missing, then the assessment is classified as post-baseline if ECG collection time is later than dosing time.
- if dosing time is missing, the assessment is classified as post-baseline.

5.3.3 Laboratory Parameters

This section provides further detail on the analysis of laboratory parameters that will be listed and summarized as described in Section 2.8.3.

Hematology

Hematologic tests include: Hemoglobin, platelets, white blood cells (WBC), red blood cells (RBC), differential (basophils, eosinophils, lymphocytes, monocytes, neutrophils (% or absolute)

The following rules will be applied to derive the WBC differential counts when only percentages are available (this is mainly for neutrophils and lymphocytes, because CTC grading is based on the absolute counts).

The method to convert the value is straightforward: for each subject, the original lab value (%) is divided by 100 and multiplied by WBC count e.g. for neutrophils (NEU):

NEU count = (WBC count) * (NEU%value/100)

In order to derive the corresponding absolute normal range, the rule to be applied depends on the availability of the % range and the absolute range for the differential:

- If absolute range NOT missing (% range is or isn't missing), then use the absolute range provided by the site
- If % range NOT missing and absolute range missing, then the % normal limits (i.e LLN and ULN) are divided by 100 and multiplied by the corresponding normal limits of WBC count, e.g. for neutrophils NEU):

  LLN for NEU count = (LLN for WBC count)* (LLN for NEU%value /100)

  ULN for NEU count = (ULN for WBC count)*(ULN for NEU%value/100)

Biochemistry

The following calculation will be applied for corrected calcium in SI unit:

Corrected calcium (mmol/L)=measured total Ca (mmol/L)+0.02 (40-serum albumin[g/L]), where 40 represents the average albumin level in g/L.

5.4 Details statistical methodology

This section provides the details of model stated in Section 2.13. Prior specifications of model parameters are provided in detail. Data analysis and decision making in real trial are illustrated using some hypothetical data scenarios. Further operating characteristics by simulation are provided in Section 3 as part of sample size justification.

5.4.1 Prior specification of model parameters

Bayesian model requires prior specification of parameters. The detailed prior specification of μ_1, μ_2,τ_1, τ_2, m_w, v_w, and p_j (j= 1,2,..,5) are described in this section.

5.4.1.1 Prior specification of exchangeability distribution (μ_1, μ_2, τ_1 and τ_2)

Prior for τ_1 and τ_2 are assumed to be half-normal distribution with scale 0.25, implying a prior 95%-interval for τ_1 and τ_2 as (0.008, 0.560), which allows for small to substantial between strata heterogeneity (see Neuenschwander Neuenschwander et al 2015).

μ_1 and μ_2 are given normal prior distributions. For first exchangeability distribution μ_1, the mean of the prior distribution ($m_{\mu 1}$) is set to logit(0.15) (or $m_{\mu 1}$= log(1/3)) which corresponds to no treatment effect. For second exchangeability distribution μ_2, the prior mean ($m_{\mu 2}$) was set to logit(0.6) (or $m_{\mu 1}$= log(3/2)). This corresponds to a substantial treatment effect. The variance parameter ($V_{\mu 1}$ and $V_{\mu 2}$) are derived using the following formula from law of total variance

$$V_{\mu i} = V(\theta) - E(\tau_i^2) \text{ and } V(\theta)=1/\pi + 1/(1- \pi) \text{ ; } \pi= 0.15, 0.60 \text{ and } i=1,2$$

This yields $V_{\mu 1}$= 7.780 ($\approx 2.789^2$) and $V_{\mu 2}$= 4.101 ($\approx 2.025^2$). This allows a considerable uncertainty on prior belief of θ.

5.4.1.2 Prior specification for stratified or "non-exchangeability" distributions (m_w and v_w)

The strata-specific normal priors for the stratified or "non-exchangeable" case are defined by m_w and v_w were. The prior median for the response probability was set as 10% (no treatment effect) i.e., m_w=logit(0.1) (or m_w = log(1/9)) and the corresponding variance (v_w) is set to 9 (=3^2) to allow large variability in prior.

5.4.1.3 Specification of mixture weights (p_j)

Finally, for each stratum j, the prior mixture weights p_j were chosen as

$$p_j = (0.25, 0.25, 0.50), \qquad j=1,\ldots,5.$$

This means that each stratum has 25% prior probability to belong to the first exchangeability distribution (μ_1 and τ_1), 25% probability to belong to the second exchangeability distribution (μ_2 and τ_2), and 50% probability to be stratified or non-exchangeable with some (or all) of the other strata.

The prior distributions are summarized in Table 5-15, which also shows the prior medians and 95%-intervals for the disease specific disease control rare (DCR) π_j.

Table 5-14 Specifications for model parameters, and prior median (95%-interval) for disease specific DCR

Parameter	Prior distribution
μ_1	N(-1.735, 2.789^2)
μ_2	N(0.405, 2.025^2)
τ_1	Half Normal(scale=0.25)
τ_2	Half Normal(scale=0.25)
m_j, v_j	N(-2.197, 3^2)
$p_j = (p_{j1},p_{j2},p_{j3})$, j=1,…,5	(0.25, 0.25, 0.5)
Prior disease specific DCR	median (2.5%, 97.5%)
π_j, j=1,…,5	0.218 (0.001, 0.981)

5.4.2 Hypothetic scenario testing

It is important to know that the design should make reasonable decisions at interim and final analysis based on the observed responses in each tumor type. This section shows on-study decisions made under the model. The hypothetical data scenarios for interim and final analysis can be found in Table 5-16 and Table 5-17 respectively. For each scenario, the probability of being "clinically meaningful" and "not clinically meaningful" (Table 2-1) are calculated by tumor type and displayed in the tables.

Table 5-16 shows 6 different scenarios for interim analyses. As stated in Section 2.13 at interim a tumor type is stopped if the probability of being "clinically meaningful" is less than 20%. At interim if no arm or tumor type shows any activity (scenarios 1 and 2) the decision based on model based inference are reasonable. The posterior probabilities of "clinically meaningful" are less than 20% for all tumor type in both scenarios (Column 6 of Table 5-16). Similarly for scenarios 3 and 4 where all tumor types show some clinically meaningful activity the proposed

decision rule suggests to "continue" all arms (posterior probability of "clinically meaningful" > 20%). The proposed design also shows reasonable decision for mixed scenarios (5 and 6). For example, under scenario 4, data for arms T3 and T5 show no clinically activity but the rest shows some activity. Based on the calculated probability of being clinically meaningful, the design allows correctly stopping (probability<0.20) for T3 and T5 at the interim while continues for the other arms.

Table 5-15 Hypothetical data scenarios and decision at interim

Scenario	Arm (Tumor Type)	No of responder/ No. of patients	Observed CBR	Posterior probability of not clinically meaningful †	Posterior probability being clinically meaningful †	Decision‡
1	T1	3/10	30%	0.8595	0.0468	Stop
	T2	1/5	20%	0.6005	0.1900	Stop
	T3	1/10	10%	0.4498	0.1878	Stop
	T4	1/10	10%	0.4685	0.1808	Stop
	T5	1/10	10%	0.4615	0.1755	Stop
2	T1	3/10	30%	0.8450	0.0550	Stop
	T2	1/5	20%	0.6223	0.1800	Stop
	T3	1/10	10%	0.4898	0.1743	Stop
	T4	1/10	10%	0.4833	0.1813	Stop
	T5	0/10	0	0.8620	0.0265	Stop
3	T1	7/10	70%	0.0843	0.7580	Continue
	T2	2/5	40%	0.1475	0.6390	Continue
	T3	3/10	30%	0.0298	0.7970	Continue
	T4	3/10	30%	0.0250	0.7980	Continue
	T5	2/10	20%	0.1128	0.6083	Continue
4	T1	8/10	80%	0.0198	0.9273	Continue
	T2	2/5	40%	0.1515	0.6040	Continue
	T3	3/10	30%	0.0253	0.7965	Continue
	T4	3/10	30%	0.0303	0.7925	Continue
	T5	2/10	20%	0.1135	0.6063	Continue
5	T1	7/10	70%	0.0373	0.8870	Continue
	T2	4/5	40%	0.0028	0.9835	Continue
	T3	1/10	10%	0.4983	0.1883	Stop
	T4	3/10	30%	0.0665	0.6865	Continue
	T5	1/10	10%	0.5123	0.1818	Stop
6	T1	7/10	70%	0.0290	0.8778	Continue
	T2	4/5	40%	0.0028	0.9870	Continue
	T3	1/10	10%	0.5538	0.1775	Stop
	T4	4/10	40%	0.0090	0.9138	Continue
	T5	0/10	0	0.9118	0.0168	Stop

† Calculated using a Bayesian hierarchical model mentioned in Section 2.13.

‡ A tumor type will be **stopped** if posterior probability being clinically meaningful is less than 20%.

Similar to interim Table 5-17 shows 5 different hypothetic scenarios for final analyses in order to illustrate final decision making process in the proposed design. As stated in Section 2.1 at final a Proof of Concept (PoC) about treatment with ceritinib will be declared for an arm (tumor type) if both of the following conditions are met:

a. Observed DCR ≥ "Disease Control Rate" threshold (Table 2-1)
b. Posterior probability of "not being clinically meaningful" (Table 2-1) is less than 20%

If no arm shows any significant activity (scenario 1) at final analysis the decision using model based inference are reasonable (declared fail to support PoC for all arms). The posterior probabilities of "not clinically meaningful" are less than 20% for all arms in this scenario (Column 5 of Table 5-17) but the observed CBR's do not cross "clinically meaningful" threshold (as stated in Table 2-1). Similarly for scenario 2 where all arms show clinically meaningful activity the proposed decision rule (to declared success) leads to PoC for all arms (posterior probability of "not clinically meaningful" < 20% and observed CBR's are more than "clinically meaningful" threshold). The proposed design also shows reasonable decision for mixed scenarios (3, 4, and 5). For example, under scenario 4, data for arms T1 and T3 show no clinically activity but the rest of the arms show activity. Based on the posterior probability of being not clinically meaningful and observed CBR, the design correctly declared fail for T1 and T3 while success for the other arms.

Table 5-16 Hypothetical data scenarios and decision at final

Scenario	Arm (Tumor Type)	No of responder/ No. of patients	CBR at final	Posterior probability of not clinically meaningful †	Posterior probability being clinically meaningful †	Decision‡
1	T1	8/20	40%	0.6133	0.1283	Fail
	T2	3/15	20%	0.6155	0.1330	Fail
	T3	2/20	10%	0.4678	0.0958	Fail
	T4	2/20	10%	0.4588	0.0968	Fail
	T5	2/20	10%	0.4670	0.0980	Fail
2	T1	14/20	70%	0.0125	0.9318	Success
	T2	6/15	40%	0.0550	0.6570	Success
	T3	6/20	30%	0.0048	0.8395	Success
	T4	4/20	20%	0.0500	0.6040	Success
	T5	4/20	20%	0.0500	0.6153	Success
3	T1	6/20	30%	0.9113	0.0113	Fail
	T2	415	40%	0.3198	0.2733	Fail
	T3	2/20	10%	0.4145	0.1703	Fail
	T4	5/20	25%	0.0255	0.6825	Success
	T5	1/20	5%	0.7243	0.0490	Fail
4	T1	8/20	40%	0.5765	0.1283	Fail
	T2	9/15	60%	0.0013	0.9800	Success
	T3	1/20	5%	0.7740	0.0403	Fail
	T4	6/20	30%	0.0040	0.8988	Success
	T5	5/20	25%	0.0205	0.7753	Success

Scenario	Arm (Tumor Type)	No of responder/ No. of patients	CBR at final	Posterior probability of not clinically meaningful †	Posterior probability being clinically meaningful †	Decision‡
5	T1	14/20	70%	0.0063	0.9500	Success
	T2	9/15	60%	0.0003	0.9810	Success
	T3	6/20	30%	0.0100	0.7983	Success
	T4	4/20	20%	0.0950	0.4938	Success
	T5	2/20	10%	0.4188	0.1608	Fail

† Calculated using a Bayesian hierarchical model mentioned in Section 2.13.

‡ A tumor type will be declared "**Success**" if posterior probability of being not clinically meaningful is less than 20% and observed CBR is at least clinically meaningful threshold.

5.4.3 Kaplan-Meier estimates

To analyze time to event variables (DOR, TTR and PFS) an estimate of the survival function will be constructed using ***Kaplan-Meier (product-limit) method*** as implemented in PROC LIFETEST with METHOD=KM option. The PROC LIFETEST statement will use the option CONFTYPE=LOGLOG. Median survival for each treatment group will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the method of [Brookmeyer and Crowley 1982]. Kaplan-Meier estimates of the survival function with 95% confidence intervals at specific time points will be summarized. The standard error of the Kaplan-Meier estimate will be calculated using Greenwood's formula [Collett 1994].

5.4.4 Confidence interval and p-value for response rate

This sections provides the programming details for the secondary objective to assess the antitumor activity of ceritinib as measured by ORR. For the statistical details of the primary objective, to assess the antitumor activity of ceritinib as measured by DCR, it is referred to Section 5.4

ORR will be summarized in terms of percentage rates with 95% CIs. An exact binomial confidence interval (implemented using SAS procedure FREQ with EXACT statement for one-way tables) will be calculated (Clopper & Pearson,1934).

SAS procedure FREQ will be used to estimate the proportion of responders (binary outcome = 1 or "Yes"), along with the associated 95% (=100 × (1 – *two-sided alpha level*)) two-sided Pearson-Clopper CI.

6 Reference

Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time, Biometrics 38, 29 - 41.

Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika; 26, 404-413.

Collet D (1994). Modelling survival data in medical research. London, Chapman & Hall.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group (2001). The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels.

FDA (2007). Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, US Department of Health and Human Services.

Hollen, P. J., Gralla, R. J., & Kris, M. G. (1995). Lung Cancer Symptom Scale (LCSS): Administration, Scoring, and Procedures Manual.

Kalbfleisch JD, Prentice RL. (2002) The Statistical Analysis of Failure Time Data. Wiley Series in Probability and Statistics.

Macdonald DR, Cascino TL, Schold SC et-al. (1990). Response criteria for phase II studies of supratentorial malignant glioma. J. Clin. Oncol. 1990;8 (7): 1277-80.

Neuenschwander B, Wandel S, Roychoudhury S, and Bailey S (2015). Robust exchangeability designs for early phase clinical trials with multiple strata. Pharmaceutical Statistics. Article first published online: 18 DEC 2015. DOI: 10.1002/pst.1730.

Wen PY, Macdonald DR, Reardon DA et-al. (2010). Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. J. Clin. Oncol. 2010;28 (11): 1963-72.