NOVARTIS

Clinical Development & Medical Affairs

LDK378

Protocol CLDK378AUS23 / NCT02186821

Modular phase II study to link targeted therapy to patients with pathway activated tumors:
Module 7 – Ceritinib (LDK378) for patients whose tumors have aberrations in ALK or ROS1

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted]

Document type: RAP Documentation

Document status: Final v2.0

Release date: 09Apr2017

Number of pages: 23

Document History – Changes compared to previous version of RAP module 3.

Version	Date	Changes

Table of Contents

1 Statistical methods planned in the protocol and determination of sample size

The primary efficacy variable, clinical benefit rate, will be analyzed by a Novartis designated Contract Research Organization (CRO), [redacted] Analysis details are provided in Appendix Q of the protocol, which is also included in Section 10 of this Report Analysis Plan (RAP).

[redacted]

All other data will be analyzed by a different Novartis-designated CRO, [redacted] according to Section 10 of the study protocol which will be available in Appendix 16.1.1 of the Clinical Study Report (CSR). Important information is given in the following sections and details will be provided, as applicable, in Appendix 16.1.9 of the CSR. All statistical analyses will be performed using SAS® Version 9.3 (or higher).

2 Statistical and analytical plans

Data from all centers that participate in this protocol will be used. The analyses stated in this document will be based on all subjects' data upon data base lock. The study is early terminated by sponsor. Last subject's survival follow-up was closed on August 4, 2016.

All statistical analyses presented in this document are related to subject background information, efficacy, and safety variables collected for the study. All analyses are descriptive and no hypothesis testing is planned (except for the primary efficacy variable, which is described in Section 10). Data from all centers will be combined for any analysis. In general, there is no stratification factor being considered for the analysis with the exception of tumor type in the primary efficacy variable analyses. Missing data will not be imputed unless otherwise specified. All data collected in the study will be presented in the listings. In general, this document will detail the methods used in summaries/tabulations.

The study population consists of adult subjects with a diagnosis of a solid tumor or hematological malignancy that have been pre-identified as having ALK or ROS1 positive mutations, translocations, rearrangements or amplifications. Pre-identification of the mutation or activation in the pathway will be performed locally at a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory prior to participation in the study. Subject must have archival or fresh tissue available for submission to a central laboratory selected by Novartis to allow for molecular testing related to pathway regulation. If the tissue is not available or is insufficient, the subject must be willing to undergo a fresh tumor biopsy to allow for this analysis. Mutation types from both local and central laboratories will be analyzed.

To be eligible for enrollment to this study, subjects must have received at least one prior treatment for their recurrent, metastatic and/or locally advanced disease and have no remaining standard therapy options anticipated to result in a durable response. Subjects must have progressive and measurable disease (per RECIST 1.1 or appropriate hematological response criteria) and are in need of treatment.

Subjects are administered with ceritinib at 750 mg once daily. A complete treatment cycle is defined as 28 days. Subjects may continue on study treatment until disease progression, withdrawal of consent, unacceptable toxicity, death, or any other reasons (e.g., physician's decision). After discontinuation of study treatment, subjects, regardless of reason for study treatment discontinuation will be followed for safety for 30 days after the last dose. Survival information will be collected every 3 months until 2 years after the last subject has enrolled in the study regardless of treatment discontinuation reason (except if consent is withdrawn). Additional survival assessments may be performed outside the 3 months follow-up schedules if a survival update is required for an interim assessment to meet safety or regulatory needs. Note that, survival follow-up has been terminated for this study on August 4, 2016.

The primary efficacy objectives of the study, clinical benefit (definition details see Section 6.1) associated with ceritinib treatment based on local investigator assessment, will be summarized. The clinical benefit rate will be analyzed using a Bayesian hierarchical model to evaluate trial success and futility, and the relationship between relevant pathway regulation and the response to treatment with ceritinib (if applicable). Analysis details are documented in Section 10.

Duration of response (DOR) based on local investigator assessment will be analyzed. In the event of a low response rate, DOR will be listed only without frequency summaries. Overall response (OR) will be summarized with the number and percentages of subjects for each response category (e.g., for solid tumor - CR, PR, SD, PD, Not evaluable). Progression-free survival (PFS) will be summarized using Kaplan-Meier (KM) methodology. Due to limited survival follow-up for the study, overall survival will only be descriptively summarized, ie, only the frequency of deaths will be provided.

Subject demographics, disease characteristics (including mutation analyses), ceritinib treatment, and safety variables will be summarized descriptively, by nominal visit (if applicable), and/or overall.

Data used for the analyses specified in this document will come from the Electronic Data Capture (EDC) system. In addition, the mutation data from the central laboratory will be collected in an Excel spreadsheet format and subsequently be uploaded to a secure FTP site by Novartis.

2.1 Change in planned analyses from the protocol

The protocol specifies that analyses are to be conducted by "patient group" which is formed by the types of cancer at study entry. That is, subjects with the same tumor type will be considered as in the same "patient group". For efficacy analyses, with exception of the primary analysis of clinical benefit rate (i.e., summarized by 'patient group'), other descriptive summaries in general will be produced overall for all ceritinib-treated subjects. Separate summaries may be produced for those groups with at least 10 subjects. All subjects will be combined for any safety analyses.

In the event of a low response rate, DOR will only be listed, and not tabulated. Subjects who had no disease progression/death will be censored at last adequate tumor assessment on or before the discontinuation/completion of the study.

For determination of PFS events, due to tumor assessments post study treatment are not collected, deaths occurred more than 30 days post treatment without prior progressive disease/relapse of disease will not be considered as an event.

For overall survival analysis, no KM method will be used due to early study termination. Descriptive summaries will be provided instead.

For vital signs, values will be descriptively summarized by nominal visit. Separate summaries of significant abnormalities will not be provided, but will be reflected in adverse events (AEs).

For cardiac imaging, due to data collection is only as clinically indicated, thus only listing will be provided.

3 Subjects and treatments

3.1 Analysis sets

The following analysis data sets will be used in the analyses:

Full analysis set (FAS): The FAS will include all subjects who have received at least one dose of the study drug. The FAS will be the primary set for efficacy analyses.

Safety set (SS): The SS will include all subjects who have received at least one dose of the study treatment and had at least one post-baseline safety assessment. A subjects who has received a dose of study treatment and who has no post-treatment safety data of any kind (i.e., no AE assessment, no vitals, no ECG, no cardiac imaging, no ECOG performance status, and no laboratory assessment) will be excluded from the SS. Note that the statement ‘no adverse event’ is considered as an assessment of AE.

3.2 Subject disposition

The number and percentage of subjects who completed study treatment or prematurely discontinued from the study treatment will be summarized with reasons for premature discontinuation. The reasons for study treatment discontinuation are as follows:

1. Adverse events
2. Lost to follow-up
3. Death
4. Disease progression
    a. Radiological assessment
    b. Clinical assessment
5. Protocol deviation
6. Non-compliance with study treatment
7. Pregnancy
8. Physician's decision
9. Subject/guardian decision
10. Study terminated by sponsor

In addition, the number and percentage of subjects who consent to be followed for survival will be presented.

For subjects who were screened but did not take the study treatment, a listing with reasons for screen failure will be provided. The reasons collected for screen failure are as follows:

1. Unacceptable past medical history/concomitant diagnosis
2. Intercurrent medical event
3. Unacceptable laboratory value(s)
4. Unacceptable test procedure result(s)
5. Did not meet diagnostic/severity criteria
6. Unacceptable use of excluded medications/therapies
7. Subject withdrew consent
8. Unknown
9. Other

In addition, a listing of subjects' deviations/violations during the study will be provided.

3.3 Study Treatment and other concomitant medications

Subjects receive ceritinib at 750 mg once daily. A complete treatment cycle is 28 days. There is no break between dosing cycles. For subjects who are unable to tolerate the protocol-specified dosing schedule, dose adjustments are permitted in order to allow these subjects to continue study treatment. Protocol section 6.3 provides the details.

In general, concomitant medications and therapies deemed necessary for the care of the subject are permitted (see Protocol Section 6.4.1 and 6.4.2), except as specifically prohibited

(see Protocol Section 6.4.3). All medications, procedures and significant non-drug therapies (including physical therapy, oxygen and blood transfusions) administered within 30 days prior to the administration of ceritinib and through 30 days after the last dose of study treatment will be recorded in the corresponding CRF page.

Detailed summaries of study treatment and other concomitant medications are described in Section 5 below.

4 Patient demographics and other baseline characteristics

Demographic data, medical history, disease characteristics and other baseline characteristics are collected at the screening visit.

In general, the baseline value will be considered as the last measurement observed prior to the first dose of study treatment, and includes assessments taken on the date of the first dose of study treatment.

Unless specified otherwise, for continuous data, the mean, standard deviation, median, 25th and 75th percentiles, minimum, and maximum will be presented. For categorical data, frequencies and percentages will be presented.

Demographic characteristics

Demographic variables include:

- Age (years), age group (<65, ≥65-<75, ≥75 years)
- Sex (Male, Female – able to bear children, premenarche, post menopausal, sterile of child bearing age)
- Race (Caucasian, Black, Asian, Native American, Pacific Islander, Other)

Medical history

Number and percentage of subjects with relevant and/or current medical history/conditions will be summarized by categories of conditions. For ongoing conditions, grade of condition (i.e., 1-4) will also be summarized. Medical history/current medical conditions will be coded using Medical Dictionary for Regulatory Activities (MedDRA). The MedDRA version used for reporting the study will be specified as a footnote in the tables/listings.

Disease characteristics

Disease characteristics variables include:

- Gene mutations as detected by local and central labs in pre-defined per-protocol mutation categories
- Primary tumor type (**Sponsor adjudicated**)

  Note: The tumor type adjudication is made based on the combination of the primary site of tumor with the tumor histology/cytology as collected in the CRFs.
- Time (in months) from initial diagnosis to the first dose of ceritinib

- Time (in months) from date of most recent recurrence/relapse to the first dose of ceritinib
- Stage at initial diagnosis
- Age of biopsy samples (in months) calculated from date of biopsy sample date to the first dose of ceritinib

Other baseline assessments

- Vital signs (height, weight, body temperature, respiratory rate, sitting pulse, sitting systolic/diastolic blood pressures [SBP/DBP])
- 12 lead ECG (QTcF, heart rate, PR duration, QT duration, QRS duration)
- Cardiac imaging (LVEF)
- ECOG performance status
- Prior lines of antineoplastic medication therapy
- Prior antineoplastic radiotherapy (yes/no)
- Prior antineoplastic surgery (yes/no)

5 Study treatment exposure and other medications

5.1 Study treatment

Duration of ceritinib treatment (in months) and actual total dose (sum of daily ceritinib treatment) will be summarized with mean, standard deviation, median, 25th percentile, 75th percentile, minimum and maximum. Duration of the treatment is defined as the last dose date of ceritinib minus the first dose date of ceritinib + 1. The number and percentage of subjects with ceritinib exposure for the following categories will be presented as well: day 1 – day 61 (≤ 2 month), day 62 – day 122 (> 2 months - ≤ 4 months), day 123 – day 183 (> 4 months - ≤ 6 months), day 184 – day 244 (> 6 months - ≤ 8 months), day 245 – day 365 (> 8 months - ≤ 12 months), and ≥ day 366 (> 12 months).

Actual dose intensity and relative dose intensity of ceritinib will be summarized. Dose intensity is defined as the total dose of ceritinib (mg) divided by duration of ceritinib (in days). Relative dose intensity (in percent) is the actual dose intensity divided by planned dose intensity (i.e., 750 mg/day), and expressed in percent.

In addition, the number and percentage of subjects who had a dose change will be summarized with reasons for the change (as per protocol: AE, dosing error, lab test abnormality, scheduling conflict, or dispensing error).

5.2 Prior and concomitant medications/non-drug therapies

Prior and concomitant medications will be coded according to the World Health Organization-Drug Dictionary (WHO-DD). The WHO-DD version will be specified as a footnote in the tables/listings. Prior and concomitant medications are mutually exclusive, as defined below:

- Prior medications are defined as any medications with an end date prior to the first dose of ceritinib;

- Concomitant medications are defined as any medications taken during the treatment period (from the first dose date and on). Prior medications that are 'ongoing' at the time of first study treatment or whose end date is after first study treatment will be considered concomitant medication.

The number and percentage of subjects with prior medications, concomitant medications that started prior to ceritinib initiation, as well as concomitant medications that started after ceritinib initiation, will be summarized by preferred term. Non-drug therapies (surgical and medical procedures) will be listed for each subject. A standard imputation method will be used for partial start/stop date of medications.

6 Efficacy and PRO evaluation

6.1 Clinical Benefit Rate (CBR)

Clinical benefit is determined by investigator assessment for each tumor assessment and is defined as responses of CR or PR or SD $\geq$ 16 weeks. In addition to tumor-specific measurements (e.g., radiological assessment of tumor response for solid tumors), progressive disease will also be considered when it was indicated either as a reason based on clinical assessment for the study treatment discontinuation or as a cause of on-treatment death. CR and PR (for solid tumors) require a confirmation that is at least a minimum of 4 weeks after the initial observation of response. Subjects who had a confirmed CR/PR or SD prior to 16 weeks, but discontinued prior to 16 weeks for reasons other than progressive disease, will have their clinical benefit defined as non-evaluable; subjects who had a confirmed CR/PR or SD prior to 16 weeks, but progressed prior to 16 weeks, will be considered as not achieving clinical benefit; subjects who had CR/PR/SD occurred prior to 16 weeks, but progressed at or after 16 weeks without evidence of CR/PR/SD at or after 16 weeks, will also be considered as not achieving clinical benefit. CBR will be analyzed by comparing achieved CBR with a historical control rate of each tumor type, and if there is at least 90% probability that the response rate in a tumor type exceeds the historical rate, then the tumor type will be considered a success; it will be considered as promising if the probability is between 80% and 90%. Details for the analyses of CBR are located in Section 10.

6.2 Overall Response Rate (ORR)

Overall response is determined by investigator assessment for each tumor assessment in the study. For subjects with solid tumors (other than lymphoma) the assessment criteria will be RECIST 1.1 and will include responses of CR and/or PR; for lymphoma, Cheson criteria will be used; for hematologic tumors, other appropriate hematological response criteria will apply. The number and percentage of subjects for different categories of overall response (e.g., for solid tumors – CR, PR, SD, PD, Not Evaluable) will be provided for solid tumors (non-lymphoma), lymphoma (if applicable), and each hematological tumor type (if applicable). Ninety-five percent (95%) exact confidence intervals (CIs) will be provided for the response rate(s) (e.g., for solid tumors – CR and/or PR) as well. For cohorts with at least 10 subjects,

overall responses will be summarized separately. Components of tumor assessments (e.g., for solid tumors – target lesion response, non-target lesion response, new lesion [yes/no]) will also be listed. In addition, Cancer Antigen-125 (CA-125) (if applicable) in the assessment of ovarian cancer response, Prostate-specific Antigen (PSA) (if applicable) in the assessment of prostate cancer will be listed.

The above analyses will be performed for the FAS.

6.3 Progression Free Survival (PFS)

PFS is defined as the time from the date of first dose of ceritinib to the date of the first documented disease progression/relapse or death due to any cause within 30 days of the last dose, and is calculated as PFS = (date of progression/relapse or death – date of first study treatment + 1). Disease progression noted as the reason for discontinuation of treatment will be considered as a progression, thus will be considered as an event for PFS.

Subjects who discontinue or complete study treatment without disease progression will be censored at last adequate tumor assessment on or before the discontinuation/completion of the study. If there is no tumor assessment for the subject, the PFS will be censored at Day 1.

PFS will be summarized and graphed using the Kaplan-Meier product-limit method[1]. The estimated median survival time, the corresponding 95% CIs[2], and 25th and 75th percentiles will be provided. In addition, survival rate estimates at 1, 2, 3, 4, 5, 6, and 12 months with 95% CIs will also be provided. For cohorts with at least 10 subjects, separate PFS analyses will be conducted for each cohort.

The above analyses will be performed for the FAS.

6.4 Overall Survival (OS)

Due to limited overall survival follow-up, only the number and percentage of all deaths will be presented.

6.5 Clinical Benefit Rate by Gene Mutations

The concordance between gene mutations detected by local and central labs will be summarized on protocol pre-defined mutations. A frequency summary will be provided for summarizing clinical benefit rate per gene mutation using mutation data provided by the central lab (on protocol pre-defined mutations only) as well as by the local lab. The data will also be presented in a listing. This analysis will be performed for the FAS.

7 Safety evaluation

7.1 Adverse events

Adverse events data will be collected at each visit. They will be assessed according to the most current version of the Common Toxicity Criteria for Adverse Events (CTCAE). If CTCAE grading does not exist for an adverse event, the severity of mild, moderate, severe,

and life-threatening, or grades 1-4 were used. CTCAE grade 5 (death) was not used in this study; rather, this information was collected in the End of Treatment/Study or Survival Information CRF page.

All AEs will be coded by Preferred Term (PT) and associated primary System Organ Class (SOC) using MedDRA (Medical Dictionary for Regulatory Activities, version 19.0 or above). The CRF collects the verbatim term for MedDRA coding.

Summaries of AE will be based on treatment emergent adverse events (TEAEs), which is defined as all new or worsening events collected from first study treatment date to last treatment date + 30. Novartis standard partial date imputation methods will be used for imputation of partial AE onset/resolution dates.

Analysis of Adverse Events

In general, the subject incidence of TEAEs will be summarized for all tables, unless otherwise noted, by SOC and PT. TEAEs will be shown in alphabetical order for SOCs and in decreasing frequencies for PTs within each SOC. Subject frequencies and percentages will be presented.

All summaries will be based on subject counts (i.e., the number and percentage of subjects who satisfy the criteria), and will depict all AEs and all grade 3/4 AEs at minimum. Each subject will be counted at most once per category. For subjects with multiple grades for the same event, only the worst grade will be counted.

The following summary tables will be provided for the SS:

- Summary of TEAEs by SOC, PT, and maximum severity
- Summary of serious TEAEs by SOC and PT
- Summary of the most common TEAEs (10% or more of subjects) by PT and decreasing frequency
- Summary of suspected TEAEs by SOC and PT
- Summary of TEAEs associated with permanent discontinuation of study drug
- Summary of TEAEs associated with dosage adjusted/temporarily interrupted of study drug
- Summary of adverse events of special interest (see details below in Section 7.1.1)

Individual records of AEs will also be listed alphabetically by the SOC/PT and then in chronological order (i.e., by start date of the event as recorded in the CRF). The listing will include subject identifier, age, race, gender, and all related event status information (e.g., grade, relationship to study medication, etc.). For the overall AE listing, AEs that are non-treatment emergent will be included. Separate AE listings that include only TEAEs (e.g., suspected AEs, serious TEAEs) will also be presented.

7.1.1 Specific safety event categories

Specific Safety Event Categories (SEC) consists of adverse events for which there is a specific clinical interest in connection with ceritinib treatment (i.e. where ceritinib may

influence a common mechanism of action responsible for triggering them) or adverse events which are similar in nature (although not identical).

Each of these SECs uses MedDRA categories (Standard MedDRA Query [SMQ], Novartis MedDRA Query [NMQ], High Level Group Term [HLGT], etc.) to group preferred terms for which there is a specific clinical interest. One SEC can be defined by one or several MedDRA categories. These SECs are defined in the current version of the ceritinib Safety profiling plan/ceritinib Case Retrieval Strategy (CRS) document.

Standard data of analysis for all SECs will be conducted as follows:

- The number and percentage of subjects will be reported by SEC, by MedDRA category and by preferred term.
- All AEs of subjects having at least one AE in that SEC will be listed.

7.2 Deaths

A subject listing of deaths with recorded principal cause of death (if available) will be provided.

7.3 Vital signs

Actual raw values and changes from baseline for vital sign results (sitting systolic and diastolic blood pressure, sitting pulse, body temperature, respiratory rate, and weight) obtained for each scheduled visit will be summarized using descriptive statistics (number of subjects, mean, standard deviation, median, 25^{th} percentile, 75^{th} percentile, minimum, and maximum). Baseline is defined as the last non-missing measurement prior to the first ceritinib dose. Vital signs will also be presented in a listing. These analyses will be performed using the SS.

7.4 Laboratory evaluations

Laboratory assessment will be collected at screening and post baseline (at scheduled visits outlined in the protocol or as clinically indicated). The collected laboratory values will be converted into standard units and the severity grade will be calculated using CTCAE v4.03. A severity grade of 0 will be assigned when the value is within normal limits. For laboratory parameters where severity grades are determined both through normal limits and absolute cut-offs, in the unlikely case that a local laboratory normal range overlaps into the higher (i.e., non-zero) CTCAE grade, the laboratory value will still be taken as within normal limits and assigned a CTCAE grade of zero. Baseline is defined as the last non-missing measurement on or prior to the first ceritinib treatment. On-treatment laboratory evaluations are defined as laboratory values collected post baseline and no later than 30 days after the last study treatment.

Abnormal on-treatment laboratory values will be summarized using shift tables (from baseline to any abnormal laboratory value post baseline) by each laboratory parameter at its worst severity grade. The number and percentage of subjects with abnormal laboratory values will be presented by CTCAE grade for each laboratory parameter that is CTCAE gradable. For laboratory parameters where CTCAE grades are not defined, shift tables using the

low/normal/high/(low and high) classification to compare baseline to the worst on-treatment value will be provided. Listings of all laboratory values (Blood chemistry, Hematology/Coagulation parameters separately) in chronologic order will be provided for each laboratory parameter. Laboratory values collected more than 30 days after study treatment discontinuation will be flagged in the listing. A separate listing will also be provided for notable laboratory abnormalities (i.e., grade 3 or 4 laboratory toxicities post start of study treatment). In addition, other labs such as urinalysis and pregnancy tests will be listed separately. These presentations will be performed for the SS.

7.5 Other safety evaluations

ECOG performance status will be summarized by visit for SS. The number and percentage of subjects in each score (0-5) will be displayed. In addition, the shift table of ECOG value from baseline to the worst score post baseline will be presented.

ECG values will be summarized descriptively by visit for the SS. Change from baseline for each collected parameter will also be summarized. The ECG parameters to be summarized include: QTcF, HR, PR, QT, and QRS.

Cardiac imaging evaluations will only be listed.

Baseline for the safety evaluation is defined as the last non-missing measurement on or prior to the first ceritinib dose.

8 Interim analyses

Scheduled interim analysis is planned for the primary endpoint of clinical benefit rate only as required by the Bayesian Hierarchical design. The first interim analysis performs after the first 30 patients have been dosed for at least 16 weeks or discontinued then every 13 weeks thereafter until study enrollment closure. After that, one CBR analysis will be done after database lock. Interim analyses may be performed more frequently dependent on enrollment rate to avoid over-enrollment in any of the disease groups. At each interim analysis, the groups will be evaluated for early futility and early success by comparing posterior quantities for the response rate to pre-specified early stopping criteria.

There is no plan for a formal interim analysis of safety or other secondary endpoints for this study. However, for publication or other purposes, interim data review of clean data will be performed as necessary. At these interim reviews, patient demographics/baseline characteristics, the primary and secondary endpoints as applicable, and all important safety endpoints will be summarized. No formal report will be issued for these interim data reviews.

9 Determination of sample size

The sample size was chosen by the usual criteria of obtaining adequate power for the alternative hypothesis of interest as shown in the protocol Tables 4.2 of Appendix Q (same as Table 2 as shown below in Section 10.4.2). This hypothesis corresponds to a generally effective treatment across groups and incorporates variation in treatment effects to reflect the realistic expectation that treatment effects may differ by group. In this setting, analytical power calculations are not possible, but the design was simulated to obtain the power of the

study as shown in Section 10 below. The sample size shown (minimum of 10 for futility stopping, minimum of 15 for early success and maximum of 30 as group cap) achieve adequate power for the alternative hypothesis. The simulations included the expected variable accrual by simulating a Poisson process with expected accrual also shown in the protocol Appendix Q (same as Section 10 below).

Given the exploratory nature of the study design, there is no power consideration for secondary objectives.

10 Bayesian Adaptive Design

10.1 Introduction

This document outlines the adaptive design for Novartis's P2P trial of LDK378.

This trial will enroll patients from at least 6 histologic groups:

- Breast
- Colorectal
- B-Cell Lymphoma
- Melanoma
- ALCL
- Ovarian

In addition, another group currently simply referred to as "Other" may be identified and enrolled throughout the study (e.g. "Other" will be a focused histologic group, but may not be explicitly identified until the trial is ongoing due to recruitment issues). We discuss the logistics in more detail below, but a decision will only be made for a group other than the 6 listed if 1) more than 3 patients are enrolled in the group, and 2) a reasonable estimate of the clinical benefit rate is available.

The primary endpoint is clinical benefit rate (CBR) in each group, with clinical benefit being assessed at 16 weeks. All patients will be assigned LDK378.

10.1.1 Primary Analysis

We let Y_i be the response indicator for the i^{th} subject, and let R_g be the assumed probability of response within a control population and $\pi_g = \Pr(Y_i = 1 \mid g_i = g)$ be the underlying probability of response for group g within the treatment group. We transform to the logit scale for modeling purposes. Let θ_g be the mean log odds treatment effect, i.e.:

$$\theta_g = \log\left(\frac{\pi_g}{1-\pi_g}\right) - \log\left(\frac{R_g}{1-R_g}\right).$$

Thus, θ_g is the group specific logistic regression coefficient for the treatment within group g. The primary analysis is a set of group specific tests that $\theta_g>0$, meaning that the treatment is better than the assumed control rate within that group. Thus, we wish to test the set of hypotheses

$$H0g : \theta g \leq 0$$

$$H1g : \theta g > 0$$

We proceed in a Bayesian fashion, assigning a prior distribution (discussed below) and computing the posterior probability of H_{1g} within each group g. If, at the final analysis,

$$\Pr(\theta g>0 \mid \text{data}) > 0.90$$

Then group g will be declared a success (thus, the final analysis produces a separate decision for each group). The trial also allows for early stopping of groups, described below.

10.1.2 Trial Logistics

The trial will enroll all available subjects in all groups for 2 years unless a group cap is reached or a group is stopped early. The trial will enroll no more than 30 evaluable subjects in each group. Interim monitoring will be conducted starting after the first 30 patients enrolled overall (across all groups) have been dosed for at least 16 weeks or discontinued, then continuing each 13 weeks thereafter. At each interim analysis, response information for the various groups will be evaluated to determine the current $\Pr(\theta_g>0 \mid \text{data})$ within each group, with sufficiently high/low values used to stop the group for success/futility. A minimum of 10 patients will be required in a group before it may discontinue enrollment for futility, and a minimum of 15 patients are required before discontinuing a group for efficacy. If a group stops enrolling early, the remaining groups will continue until the end of 2 years or until the other groups reach their own early stopping criteria. The final analysis will occur after 2 years of trial duration, with all subjects enrolled at 2 years followed up to their endpoint.

The trial will enroll subjects in all listed groups. In addition, should other groups be identified throughout the trial, the following mechanism will be used. If another group is identified, it will not be placed into the statistical analysis unless 3 subjects enroll within the group (thus, the trial may enroll multiple possible groups within the "other" category, but a group will only be added to the list if at least 3 patients enroll from that group. Thus, it is possible (but not viewed as likely) that multiple "other" groups may be added to the trial if the trial has sufficient enrollment in multiple "other" groups. In addition to sufficient enrollment, the sponsor must have a reasonable estimate of the control clinical benefit rate.

Subjects within the "other" category which do not reach the minimum 3 subject enrollment will be excluded from the interim and final analyses. Thus, in the statistical analysis below, the number of groups may be 4 (if no other groups meet the requirements for inclusion in the study) or more if other groups are identified throughout the course of the study. As the study continues, early interim analyses may be based on fewer groups than later interim analyses, as the interim analyses will include whatever groups have satisfied the criteria at the time of the analysis.

10.2 Statistical Modeling

We let Y_i be the response indicator for the i^{th} subject, and let R_g be the probability of response within a control population and $\pi_g = \Pr(Y_i = 1 \mid g_i = g)$ be the underlying probability of response for group g within the trial. We transform to the logit scale for modeling purposes. Let θ_g be the mean log odds treatment effect, i.e.:

$$\theta_g = \log\left(\frac{\pi_g}{1-\pi_g}\right) - \log\left(\frac{R_g}{1-R_g}\right).$$

The statistical design borrows information across groups with a hierarchical model. The hierarchical model allows dynamic borrowing of information between groups such that more borrowing occurs when the groups are consistent and less borrowing occurs when the groups differ. In this way, the model is a compromise between the two alternate extremes of either a completely pooled analysis or a separate analysis in each group. We additionally incorporate a clustering mechanism that allows borrowing within clusters but treats clusters separately. This minimizes borrowing across groups that are quite different in terms of CBR.

The purpose of such an analysis is to produce higher power or lower type I error in situations where we see some commonality (identical effects are not required) among the groups. The model will borrow more in situations where the groups appear similar than situations where the groups appear different.

10.2.1 Hierarchical Model with Clustering

Our hierarchical approach involves two stages. The goal of both stages is to allow the data to drive the amount of borrowing across groups. If the data indicate a large amount of borrowing is appropriate (due to similar results), the model will borrow more and thus increase the overall power of the trial within each group. In contrast, if the data indicate a small amount of borrowing is appropriate (due to dissimilar results) the model will adjust and each group will stand more on its own. This "dynamic" borrowing property is distinct from other approaches which use a fixed informative prior or apriori assume an amount of borrowing across groups. Here the approach includes two stages to identify the appropriate amount of borrowing based on the data.

The first stage of model places the groups into distinct clusters. The purpose of this stage is to minimize borrowing of information across groups that appear to be quite different. Thus, for example, should 2 of the groups appear similar while the others differ significantly, the model may place a large probability on two clusters, one containing the two similar groups with the other containing the remaining groups. The model incorporates the uncertainty of the data in this determination, producing a probability distribution over the possible clusterings. Thus, in our example, the model may consider it highly likely that the 2 similar groups are in one cluster with the remaining groups in another, but it would also retain lower probabilities on the possibility all groups are in one cluster (e.g. we are simply seeing differences in the two groups by chance) as well as other possibilities. The complete analysis averages over this uncertainty. This clustering approach is implemented through a Dirichlet Process Mixture (DPM) model, described in the appendix.

At the second stage, we place hierarchical models over the groups within each cluster (thus, conditional on the clustering, there is no borrowing of information across clusters, only within clusters). The hierarchical model assumes that the θ_g have an across groups distribution

$$\theta_g \sim N(\mu, \tau^2)$$

The across group mean μ and variance τ^2 are unknown, and hence have a prior distribution which is combined with the data to produce estimates of μ and τ^2.

The variance component τ controls the degree of borrowing among groups. Small values of τ result in a greater degree of borrowing while large values of τ correspond to less borrowing.

The parameter τ is estimated using the data, so the observed between group variation is a key component of the model behavior.

Combined, the two stages allow groups with similar results to borrow information between them (they will have a high probability of being in the same cluster) while groups with different results with borrow far less information between them (they will have a low probability of being in the same cluster).

Details of the two stages may be found in the appendix.

10.3 Evaluation of Trial Success and Futility

Interim monitoring will occur after the first 30 patients, then every 13 weeks thereafter. At each interim analysis, the groups will be evaluated for early futility and early success by comparing posterior quantities for the response rate to pre-specified early stopping criteria.

10.3.1 Early Futility

If there is less than 10% probability that the response rate in a group exceeds the historical rate R_g, then the group will stop enrollment early for futility. Formally, enrollment will stop early for futility if:

$$\Pr(\pi_g > R_g) < 0.10.$$

A group is only eligible for early stopping once a minimum of 10 patients has been evaluated for response in that group.

10.3.2 Early Success

If there is at least 95% probability that the response rate in a group exceeds the historical rate, then the group will stop enrollment early for success. Formally, enrollment will stop early for success if:

$$\Pr(\pi_g > R_g) > 0.95.$$

A minimum of 15 subjects will need to be evaluated prior to declaring a group to be efficacious.

10.3.3 Final Analysis

In addition, recall the final analysis will occur when both accrual and follow-up are complete in all groups. If, at the completion of the trial, there is at least 90% probability that the response rate in a group exceeds the historical rate, then the group will be considered a success. Formally:

$$\Pr(\pi_g > R_g) > 0.90.$$

10.4 Simulation

We evaluated type I error and power for each of the 7 possible groups (including one "other" category) under a variety of possible "truths" indicating various possible true underlying probabilities within each group. Note as described above the 7th "other" group is only included in the analysis if it enrolls at least 3 subjects.

10.4.1 Assumptions

Accrual - Table 1 shows the assumed two year expected accrual number as well as the assumed control rates for the 7 histologic groups (including "other"). Note these are averages, the actual number of available patients is simulated as a Poisson distribution with the specified mean. Also note that the group cap of 30 applies, and thus if the number of available patients in a group exceeds 30, only the first 30 available patients in that group will be enrolled in the study. Note given the low assumed enrollment rate for "other" (2), there a high likelihood these subjects will not enroll sufficiently to enter the main analysis.

Dropouts – We assume no dropouts for the purpose of this simulation.

Control Rates –Table 1 also shows the assumed control rates.

Histology	**Assumed 2 year accrual**	**Assumed Control Rate (R_g)**
Breast	20	0.69
Colorectal	15	0.26
B-cell Lymphoma	6	0.60
Melanoma	8	0.30
ALCL	6	0.90
Ovarian	15	0.65
Other (where applicable)	2	0.50

Table 1 – Assumed 2 year accrual rates (accrual is simulated as a Poisson process with these means) and assumed control CBR values used in the simulations.

We consider four possible scenarios, or possible "truths" in the simulation. These consisted of a null scenario (where the treatment has no effect for any group), an alternative scenario (where the treatment is effective in all groups), a scenario where the treatment was effective in two of the groups, and a scenario where the treatment was effective in half of the groups.

Treatment Rates - The treatment rates for each scenario are shown in the table below. Values identical to the control are shown in bold, while values greater than the assumed control rate are shown in italics.

	Null	**Alternative**	**Two**	**Half**
Breast	0.69	0.88	0.88	0.69
Colorectal	0.26	0.50	0.26	0.50
B-cell Lymphoma	0.60	0.82	0.60	0.60
Melanoma	0.30	0.55	0.55	0.55
ALCL	0.90	0.97	0.90	0.90
Ovarian	0.65	0.85	0.65	0.85
Other (where applicable)	0.50	0.75	0.50	0.75

Simulation Details – For each scenario we simulated 1000 trials. For each interim within each trial, we ran 50,000 MCMC iterations after a 1,000 MCMC iteration burnin.

10.4.2 Results

Table 2 below provides the probability of group success for each of the groups

Group (Exp N)	Null	Alternative	Two	Half
Breast (20)	**0.126**	*0.924*	*0.815*	**0.246**
Colorectal (15)	**0.059**	*0.813*	**0.116**	*0.743*
B-cell Lymphoma (6)	**0.063**	*0.692*	**0.125**	**0.180**
Melanoma (8)	**0.052**	*0.698*	*0.517*	*0.627*
ALCL (6)	**0.003**	*0.611*	**0.045**	**0.148**
Ovarian (15)	**0.082**	*0.839*	**0.151**	*0.771*
Other (where applicable) (2)	**0.011**	*0.509*	**0.069**	*0.294*

Entries in bold represent groups where the treatment effect is 0 (e.g. the treatment is ineffective). Thus, entries in bold are type I errors. Entries in italics appear where the treatment is effective, and thus indicate the power of the design.

Generally, type I error is controlled at 0.10 under the null scenario (the borrowing compensates for the multiple interim analyses) and power is an increasing function of the expected sample size (power for Breast, Colorectal, and Ovarian are generally high, particularly in the alternative case). In the alternative scenario there remains decently high probability of success even in the lower enrolling groups. When fewer groups are effective in truth, the scenarios "half" and "two" are harder to discern, however the model still retains around 75-80% power to isolate the highest enrolling groups. In these scenarios identifying the lower enrolling groups becomes more problematic, with power between 50-65%.

Power is reduced and type I error is inflated when the truth is a mixture of effective and ineffective treatment effects across the groups. Generally power again is a function of the sample size.

10.5 Modeling Details

Recall at the first stage the groups are clustered according to a Dirichlet Process Mixture Model.

The number of clusters is not assumed to be known in advance but will instead be inferred from the data using Dirichlet Process Mixtures (DPM). The DPM looks across all the possible clusterings of the groups and assigns a probability to each based on the data. The prior distribution in a DPM is governed by a parameter α. When α is small, the prior favors large clusters. As α tends to zero, the prior tends to place all its mass on a single cluster containing all the groups. As α increases, the prior places more mass on clusterings with a large number of clusters. As α becomes very large, the prior places all of its mass on having a separate cluster for each group (that is, no borrowing across groups). Thus, by specifying extreme values of the prior one could force the groups into one cluster or force the groups to be analyzed in separate clusters. Here we choose a moderate version of α=2 (common values might be anywhere between 0.5 and 5) and allow the data more control over the clustering.

The details of the prior are as follows. Let z_g represent the cluster to which group g belongs. Then $z_g \sim$ Categorical($\boldsymbol{p}$), where $\boldsymbol{p}$ is the vector such that p_k is the probability that a group belongs to cluster k and $\sum_{k=1}^{\infty} p_k = 1$. We construct $\boldsymbol{p}$ using a stick-breaking process:

$$p_k = \beta_k \prod_{i=1}^{k-1}(1-\beta_i)$$

and

$$\beta_k \sim \text{Beta}(1, \alpha).$$

A large value of α thus removes a very small amount of probability for $\boldsymbol{p}$, resulting in many clusters, while a small value of α tends to produce probabilities near 1 for the first cluster.

Conditional on the clustering, we fit a hierarchical model which has an across groups distribution

$$\theta_g \sim N(\mu, \tau^2)$$

As discussed above, this across groups distribution states that within a cluster we expect to see some variation in the parameters, with that variation governed by τ. When τ is small, there is minimal variation across groups within a cluster, and thus within the cluster the model would approach pooling. In contrast, when τ is large we expect large amount of across group variation, and thus even though the groups are in the same cluster the θ_g values may be quite different. Apriori we have no way of knowing τ, so we estimate it using the data combined with the prior distributions

$$\mu \sim N(0, 1.82)$$

and

$$\tau^2 \sim IG(3,0.5),$$

where *IG(α, β)* is the inverse gamma distribution defined by:

$$f(x|\alpha,\beta) = \frac{\beta^{\alpha} e^{-\beta/x}}{x^{\alpha+1}\Gamma(\alpha)}.$$

When the entire model is implemented (via Markov Chain Monte Carlo) we consider the full joint distribution of the clustering combined with the hierarchical model parameters. We average over the entire range of the uncertainty in the parameters to produce the posterior distribution for each group parameter θ_g, which is then used to drive the decisions in the model.

11 References

1 **Kaplan EL, Meier P.** Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc*. 1958; 53: 457-81.

2 **Brookmeyer R., Crowley, J.** A Confidence Interval for the Median Survival Time. *Biometrics*, 1982; 38: 29-41