NOVARTIS

Clinical Development

Ribociclib (LEE011)

Protocol CLEE011XUS03 / NCT02187783

Modular phase II study to link targeted therapy to patients with pathway activated tumors:
Module 8 – Ribociclib for patients with CDK4/6 pathway activated tumors

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted]

Document type: RAP Documentation

Document status: Final v2.0

Release date: 27-Sep-2015

Number of pages: 32

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

Version	Date	Changes

Table of Contents

1 Introduction

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 R of the protocol, which is also included in Section 8 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

2.1 Study design

This is a phase II, open label study to determine the efficacy and safety of treatment with ribociclib (LEE011) in patients with solid tumors or hematological malignancies that have been pre-identified as having CDK4/6, cyclin D1/3, or p16 aberrations and whose disease has progressed on or after standard treatment. Pre-identification of the relevant gene mutation or amplification status will be performed locally at a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory prior to participation on the trial.

To be eligible for enrollment to this study, patients must have archival tissue available for submission to allow for molecular testing related to pathway activation. If tissue is not available or is insufficient the patient must be willing to undergo a fresh tumor biopsy to allow for these analyses. Mutation types from both local and central laboratories will be analyzed.

Patients 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. Patients received ribociclib on a flat scale of 600 mg once daily, 3 weeks on and 1 week off. A complete treatment cycle is defined as 28 days. Patients may continue on study treatment until disease progression, withdrawal of consent, unacceptable toxicity, death, or any other reasons (e.g., physician's decision). The patient may not receive any additional anti-cancer therapy during treatment with ribociclib. After discontinuation of study treatment, patients, 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 patient has enrolled in the study regardless of treatment discontinuation reason (except if consent is withdrawn). If the study's primary efficacy endpoint is not met, Novartis may decide not to conduct survival follow up for the study.

2.2 General overview of statistical methodology

Data from all centers that participate in this protocol will be used. 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.

The analyses stated in this document will be based on all patients' data up to the time when all patients have had the opportunity to complete at least 4 cycles (or 16 weeks) of study treatment or discontinued study. The data cut-off date will be the date when the last patient on treatment completes 4 cycles of study treatment. This will be the cut-off point for the clinical study report (CSR). Additional data for patients continuing to receive study treatment beyond the data cut-off date for the primary CSR will be reported once all patients have discontinued treatment or been lost to follow-up.

All statistical analyses presented in this document are related to patient background information, efficacy, and safety. All analyses are descriptive and no hypothesis testing is planned (except for the primary efficacy variable, which is described in Section 8 below). For continuous data, the mean, standard deviation (SD), median, 25^{th} and 75^{th} percentiles, minimum, and maximum will be presented. For categorical data, frequencies and percentages will be presented. In general, data from all centers will be combined for any analysis. There will be no stratification factor considered for the analysis with the exception of tumor type in the primary efficacy variable analyses. Missing data will not be imputed. All data collected in the study will be presented in the listings.

The primary efficacy objectives of the study, clinical benefit (e.g., for solid tumor – CR, PR, stable disease [SD] $\geq$ 16 weeks) associated with ribociclib treatment based on local investigator assessment, will be analyzed by using a Bayesian hierarchical model for evaluation of trial success and futility, and the relationship between pathway activation and response (if applicable). Analysis details are documented in Section 8.

In addition, with all solid tumor types combined and separately for each hematological tumor type, the clinical benefit rate (CBR) will be provided with its 95% exact confidence interval (CI) (using clopper-pearson[1] method); overall response rate (ORR, e.g., partial response [PR] and complete response [CR] for solid tumor) and its 95% CI will be presented. Summary on numbers and percentages of patients for each response category (e.g., for solid tumor - CR, PR, stable disease [SD], progressive disease [PD], not evaluable [NE]) will also be provided. For tumor types that have ten (10) or more patients, separate summaries will be performed by the tumor type using the same methodology. In the event that there are fewer than 4 patients for a hematological tumor type, only listing will be provided.

Duration of response (DOR) based on local investigator assessment will also be analyzed. Progression-free survival (PFS) and overall survival (OS) will be summarized using Kaplan-Meier (KM) methodology.

Patient demographics, disease characteristics (including mutation analyses), ribociclib treatment, and safety variables will be analyzed descriptively.

Details of the above analyses are presented in the following sections.

2.3 Change in planned analyses from the protocol

Cardiac imaging is only collected post baseline where clinically indicated and the value and potential abnormality will only be listed, and not summarized.

Summaries will be conducted for all patients combined with the following exceptions –

- Primary efficacy analysis on the clinical benefit via Bayesian hierarchical model
- Other clinical benefit and tumor response summaries for tumor types that have 10 or more patients enrolled.

In the event of a low response rate, DOR will be listed only.

2.5 Definitions

2.5.1 Study drug

Study drug = ribociclib 600 mg once daily, 3 weeks on and 1 week off in a cycle of 28 days.

2.5.2 Date of first administration of study drug

The date of first administration of study drug is defined as the first date when a nonzero dose of study drug is administered and recorded on the dose administration electronic case report form (eCRF). The date of first administration of study drug will also be referred as start of study drug.

2.5.3 Date of last administration of study drug

The date of last administration of study drug is defined as the last date when a nonzero dose of study drug is administered. Last date of study drug will also be referred as end of study drug.

2.5.4 Study day

The study day, describes the day of the event or assessment date, relative to the reference start date (date of first administration of study drug).

The study day will be calculated as the difference between the date of the assessment and the start of study treatment plus 1. If the date of the assessment is before the start of study treatment the study day will be negative and will be calculated as the difference between the date of the assessment and the start of study treatment.

2.5.5 Baseline

In general, the last available assessment before or at date of start of study treatment will be used as the 'baseline' assessment. For electrocardiogram (ECG), the average of the triplicate assessments prior to the first administration of study drug will be used as the 'baseline' assessment. In the event of missing pre-dose ECG time-point on the day of study treatment start, the screening ECG assessment will be used as the baseline.

2.5.6 On treatment assessment

On-treatment assessment is defined as any assessment obtained in the time interval: Date of first administration of study treatment through the date of last administration of study treatment + 30 days. If the patient is still on treatment at the time of the cut-off, on-treatment assessments include any assessment recorded in the database up to the cut-off and which occur after the start date of study treatment.

Data listings will include all assessments, flagging those which are not on-treatment assessments.

2.5.7 Last contact date

The last contact date will be derived for patients not known to be dead at the analysis cut-off using the last complete date among the following:

- All assessment dates (e.g. vital signs assessment, performance status assessment). Note, only a true on study assessment date or patient contact date will be used. If there is a visit date without evidence of any actual assessment performed that date will not be used. No dates post cut-off will be used.
- Medication dates including study medications, and concomitant medications administered.
- Adverse event dates
- Last contact date collected on the 'Survival information' eCRF (only if the patient status is not unknown).
- The cut-off date will not be used for last contact date, unless an actual assessment or patient contact was performed.

Important note: imputed dates (e.g. the analysis cut-off date programmatically imputed to replace the missing end date of a dose administration record) will not be used to derive the last contact date, except for the following: partial date's imputation is allowed to be used for

event (death) and for censoring date only if coming from Survival Information page.

The last contact date will be used for censoring of patients in the analysis of overall survival.

2.5.8 Year, month and week

For reporting purpose below mentioned rule will be followed to convert a year, month and week in days.

1 year = 365.25 days

1 month = 30.3475 days

1 week = 7 days

3 Patients 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 patients 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 patients who have received at least one dose of the study drug and had at least one post-baseline safety assessment. Any patient who has received a dose of study drug and who has no post-treatment safety data of any kind (i.e., no AE assessment, no vital signs, no ECG, no cardiac imaging, no ECOG performance status, and no laboratory assessment) will be excluded from the SS. Note that the statement 'no AE' is considered as an assessment of AE.

The number and percentage of patients in each analysis sets will be presented.

3.2 Protocol deviations

The complete list of CSR reportable protocol deviations are listed in the data handling plan. CSR reportable protocol deviations will be listed.

3.3 Patient disposition

The FAS analysis set will be used for patient disposition summaries. Following categories will be summarized:

1. Number (%) of patients who are treated
2. Number (%) of patients who are still on treatment
3. Number (%) of patients who discontinued study drug
4. Reasons for discontinuation of study drug
5. Number (%) of patients who consent to be followed for survival follow-up

6. Number (%) of patients who lost to follow-up during survival follow-up period (Note: it is defined as the patients whose last known survival status based on the survival follow-up contact occurred more than 6 months before the cut-off date)

For patients who were screened but did not take the study treatment, a listing with reasons for screen failure will be provided.

3.4 Patient background, demographics and baseline characteristics

Demographic and baseline characteristics data, disease characteristics, medical history, and prior anti-neoplastic therapies are collected at the screening visit. Descriptive summaries and/or listings will be provided. The FAS analysis set will be used for the presentation.

Demographic and baseline characteristics

Demographic characteristic variables include:

- Age (years), age group (<65 and ≥65; <75 and ≥75 years)
- Sex (Male, Female)
- Race (White, Black or African American, Asian, Other [including American Indian or Other Alaska Native, Native Hawaiian or Other Pacific Islander, and Other])

Baseline characteristic variables include:

- Height and weight
- ECOG performance status
- Cardiac imaging (Left Ventricular Ejection Fraction [LVEF])

Disease characteristics

Disease characteristic variables include:

- 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 ribociclib
- Time (in months) from date of most recent recurrence/relapse to the first dose of ribociclib
- Presence of B symptoms (for lymphoma patients only)
- Stage at initial diagnosis
- Prior lines of antineoplastic medication therapy
- Prior lines of chemotherapy
- Prior antineoplastic radiotherapy (yes/no)
- Prior antineoplastic surgery (yes/no)

Other patient characteristics

- Gene mutations as detected by local and central labs in pre-defined per-protocol mutation categories

- Age of biopsy samples (in months) calculated from date of biopsy sample date to the first dose of ribociclib

Medical history

Medical history and ongoing conditions, including cancer-related conditions and symptoms will be summarized and listed. The summaries will be presented by primary system organ class and preferred term. Medical history/current medical conditions are coded using the Medical dictionary for regulatory activities (MedDRA) terminology (v18.0). A separate summary will be presented for ongoing medical conditions by severity of the condition (i.e., grade 1 to grade 4).

Prior anti-neoplastic therapies

Prior anti-neoplastic therapies will be listed in three separate listings:

- Medications
- Radiotherapy
- Surgery

3.5 Study Treatment and concomitant medications

The safety set will be used for all analyses associated with study treatment and other medications/non-drug therapy.

3.5.1 Study treatment

Patients receive ribociclib (600 mg) once daily, 3 weeks on and 1 week off. A complete treatment cycle is 28 days. There is no break between dosing cycles. For patients who are unable to tolerate the protocol-specified dosing schedule, dose reductions or interruptions are permitted to manage drug-related toxicities.

- When dose reduction is necessary, the dose of ribociclib may be reduced to 400 mg.
- If an additional dose reduction is required, ribociclib may be reduced to 200 mg.
- Once the ribociclib dose is reduced it cannot be re-escalated.

Protocol Section 6.3 provides the detailed information for dose modifications or dose delays.

3.5.1.1 Duration of exposure

Duration of exposure (days) for study drug is defined as the days between the last ribociclib dose and first ribociclib dose, that is to say, duration of exposure (D1) = (date of last administration of ribociclib) – (date of first administration of ribociclib) + 1. Summary statistics will be displayed for the duration of exposure.

Duration of exposure to study drug will also be categorized into time intervals (< 3 months, 3 – < 6 months, 6 – < 9 months,…, etc.). The number and percentage of patients in each category will be presented.

3.5.1.2 Adjusted duration of exposure

Adjusted duration of exposure (days) for ribociclib (D) is the number of ribociclib dosing days a patient would be expected to have received per protocol, given their duration of exposure to study treatment. Since ribociclib follows a 3 weeks on, 1 week off schedule, the adjusted duration of exposure to ribociclib is the duration of exposure to study treatment minus the planned off days for ribociclib. The adjusted duration of exposure to ribociclib (in days) is therefore 21 x (# completed 28 day cycles) + minimum of (21, duration of last incomplete cycle). Summary statistics will be displayed.

3.5.1.3 Cumulative dose and average daily dose

Cumulative dose is defined as the total dose actually given during the study treatment exposure. Average daily dose is defined as [cumulative dose (mg) / cumulative dosing days (days)].

The dosing record, as collected in the eCRF, includes planned one week drug breaks, thus cumulative dose calculation will be adjusted according to the 3 weeks on, 1 week off schedule. That is, for each dosing record, the calculated total dose of ribociclib will be the daily ribociclib dose x 21 x (# completed 28 day cycles) + daily ribociclib dose x minimum of (21, duration of last incomplete cycle). Since the eCRF does not collect cycle start dates, it will be assumed that a new cycle starts when ribociclib resumes after a temporary interruption.

Cumulative dosing days are calculated based on actual dosing days, excluding both one week drug break and temporary dose interruption.

Cumulative dose and average daily dose will be summarized using descriptive statistics.

3.5.1.4 Dose intensity and relative dose intensity

Dose intensity (DI) will be calculated as follows:

DI = Cumulative dose (mg) / Adjusted duration of exposure (days)

Planned dose intensity (PDI) is the assigned dose by unit of time planned to be given to patients as per protocol. The planned dose intensity for ribociclib during the period of time the patient received drug is 600 mg/day.

RDI = DI (mg/day) / 600 mg/day.

DI and RDI will be summarized using descriptive statistics.

3.5.1.5 Dose interruption and dose reduction

Dose interruption: An interruption is defined as a 0 mg dose given on one or more days during the treatment period where a patient is not on the "off" part of a treatment cycle. For patients who have dose interruption checked but never resume the non-zero dose, the dose interruption will not be counted.

Dose reduction: A reduction is defined as a decrease in dose from the protocol planned dose or a decrease from the previous non-zero dose, during the dosing period, even if this decrease has been directly preceded by an interruption. For example, in the sequence 600 mg – 0 mg –

400 mg, the 400 mg dose will be counted as a reduction; in the sequence 400 mg – 0 mg – 400 mg, the 400 mg dose will not be counted as a reduction.

The number and percentage of patients who have dose reductions or interruptions, and the reasons for such reductions/interruptions, will be summarized separately.

3.5.2 Concomitant medications/non-drug therapies

In general, concomitant medications and therapies deemed necessary for the care of the patient are permitted (see Protocol Section 6.4.1), except as specifically prohibited (see Protocol Section 6.4.2). All medications, procedures and significant non-drug therapies (including physical therapy and blood transfusions) administered within 30 days prior to the administration of ribociclib and through 30 days after the last dose of study drug will be recorded in the corresponding eCRF page.

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) besides the study treatment that were administered to a patient, preceding or coinciding with the study assessment period.

Concomitant medications entered into the database will be coded using the World Health Organization (WHO) Drug Reference List to allow for categorization by preferred term. In addition to categorizing medication data by preferred term, drugs will be classified according to their ATC classification in order to present and compare how they are being utilized. Non-drug therapies will be coded using the MedDRA terminology (v18.0).

Concomitant medications taken concurrently with the study drugs will be listed and summarized by ATC class, preferred term by means of frequency counts and percentages. These summaries will include medications starting on or after the start of study treatment but no later than 30 days after last dose of study treatment or medications starting prior to the start of study treatment and continuing after the start of study treatment. Similarly, significant non-drug therapies will be summarized by primary system organ class (SOC) and preferred term.

Any prior medications or significant non-drug therapies starting and ending prior to the start of study treatment or starting more than 30 days after the last date of study treatment will be listed.

The safety set will be used for all concomitant medication/non-drug therapy tables and listings.

4 Efficacy evaluation

All efficacy analyses will be performed in FAS.

4.1 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 for $\geq$ 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 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. Patients who had 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; patients who had confirmed CR/PR or SD prior to 16 weeks, but progressed prior to 16 weeks, will be considered as not achieving clinical benefit; patients 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 80% probability that the response rate in a tumor type exceeds the historical rate, then the tumor type will be considered a success. Details for the Bayesian analyses of CBR are located in Section 8.

In addition to the Bayesian analyses, with all solid tumor types combined and separately for each hematological tumor type, the CBR estimate will be provided with its 95% exact CI. For tumor types (within the solid tumors) that have ten (10) or more patients, separate summaries will be performed by the tumor type. In the event that there are fewer than 4 patients for a hematological tumor type, only listing will be provided for that tumor type.

4.2 Overall Response Rate (ORR)

Overall response is determined by investigator assessment for each tumor assessment in the study. For patients with solid tumors, the assessment criteria will be RECIST 1.1 and will include responses of CR and/or PR. For hematological tumors, other appropriate criteria (see Protocol Appendix section for corresponding criteria) will be used to determine the responses. Ninety-five percent (95%) exact CI will also be provided for the response rates. The number and percentage of patients for different categories of overall response (e.g., for solid tumor - CR, PR, SD, PD, and NE) will be presented. These analyses will be performed with all solid tumor types combined and separately for each hematological tumor type. For tumor types that have ten (10) or more patients (within the solid tumors), separate summaries will be performed by the tumor type. In the event that there are fewer than 4 patients for a hematological tumor type, only listing will be provided for that tumor type.

Components of tumor assessments (target lesion response, non-target lesion response, new lesion [yes/no]) will be listed. In addition, Cancer Antigen-125 (CA-125) in the assessment of ovarian cancer response, or Prostate-specific antigen (PSA) in the assessment of prostate cancer response will also be listed.

4.3 Duration of response

DOR applies only to patients who responded, e.g., for solid tumor, responder is defined as the patient's best overall response being CR or PR. The start date is the date of first documented response, and the end date is the date of event defined as the first documented progression/relapse or death due to any cause within 30 days of the last study drug dose date. In other words, the start date will be determined using the time the response was first determined and not using the time the response was confirmed. If a patient has not had an event, duration is censored at the date of last adequate tumor assessment. Distribution of DOR will be estimated using the Kaplan-Meier method and the median response duration will be

presented along with 95% confidence interval only if there are sufficient numbers of events, otherwise, only listing of DOR will be provided. Analyses will be based all tumor types combined.

4.4 Progression Free Survival (PFS)

PFS is defined as the time from the date of first dose of ribociclib to the date of the first documented disease progression/relapse or death due to any cause within 30 days of the last study drug dose date, 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 also be considered as a progression event for PFS analyses.

Patients 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. For patients who are still on study treatment at the time of data cut-off date for the primary analyses, PFS will be censored at the last adequate tumor assessment before the data cut-off date for the analysis of PFS for the primary CSR. Patients who do not have any post baseline tumor assessment will be considered as censored at Day 1.

PFS will be summarized and graphed using the Kaplan-Meier product-limit method[2]. The estimated median survival time, the corresponding 95% CIs[3], and 25th and 75th percentiles will be provided. In addition, survival-rate estimates with 95% CIs will also be provided at time points, such as 1, 2, 3, 4, 5, 6, 9 and 12 months. Analyses will be based on all tumor types combined. For tumor types that have ten (10) or more patients, separate summaries will be performed by the tumor type.

4.5 Overall Survival (OS)

Overall survival, defined as the time from the date of the first dose of ribociclib to the date of death due to any cause will be analyzed using methods similar to those stated above for PFS. That is, the 25th percentile, median, and 75th percentile of survival time with corresponding 95% CIs will be provided; survival curve and survival rates at time points, such as 3, 6, 9, 12, 18, and 24 months will also be presented. Patients who are alive at the time of discontinuation/completion of the study will be censored at the last contact date (as defined in Section 2.5.7). Analyses will be based on all tumor types combined. For tumor types that have ten (10) or more patients, separate summaries will be performed by the tumor type.

Handling missing month/day in date of death

For rare cases when either day is missing or both month and day are missing for the date of death, the follow imputation rules will be implemented:

- If only day is missing, then impute maximum of [(1 mmm-yyyy), minimum of (any valid date from data base used for deriving last contact date + 1, cut-off date)].

- If both day and month are missing, then impute maximum of [(1 Jan-yyyy, minimum of (any valid date from data base used for deriving last contact date + 1, cut-off date)].

4.6 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 local lab. The data will also be presented in a listing.

5 Safety evaluation

The assessment of safety will be based mainly on the frequency of adverse events and on the number of laboratory/ECG values that fall outside of pre-determined ranges. Other safety data (e.g. vital signs) will be considered as appropriate.

All safety outputs will use the safety set. The safety summary tables will only include "on-treatment" assessments (see the definition of "on-treatment" in Section 2.5.6), i.e. those collected on or after the first date of study treatment and collected no later than 30 days after the date of last study treatment administration. The AEs started before the first dose but worsened during the treatment are also considered as 'on-treatment' events. All safety assessments will be listed and those collected later than 30 days after the last treatment/exposure date will be flagged.

5.1 Adverse events

5.1.1 Coding of AEs

Adverse events are coded using the MedDRA (v18.0).

5.1.2 Grading of AEs

AEs will be assessed according to the 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 v4.03 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).

If CTCAE grading does not exist for an adverse event, grades 1 – 4 corresponding to the severity of mild, moderate, severe, and life-threatening will be used. CTCAE grade 5 (death) will not be used in this project; rather, this information will be collected on the "End of Treatment", "30-days safety follow-up" or "Survival Information" eCRF pages.

5.1.3 General rules for AE reporting

AE summaries will include all AEs starting on or after study Day 1 (i.e. on or after the day of the first intake of study treatment) and starting no later than 30 days after the last treatment/exposure date. All AEs will be listed. AEs starting prior to study Day 1 and AEs starting later than 30 days after the last treatment/exposure date will be flagged in the listings.

AEs will be summarized by presenting the number and percentage of patients having at least one AE, having at least one AE in each body system/primary system organ class, and for each preferred term using MedDRA coding. A subject with multiple occurrences of an AE will be counted only once in the AE category.

Separate AE summaries will be presented by primary system organ class (SOC), preferred term (PT), and maximum CTC grade. A patient with multiple CTC grades for an AE will be summarized under the maximum CTC grade recorded for the event. In the summaries presented by grade, all AEs will be pooled regardless of whether they are CTC gradable or not.

Any information collected (e.g. CTC grades, relatedness to study drug, action taken etc.) will be listed as appropriate.

5.1.4 AE summaries

The following summary tables will be provided:

- AEs, regardless of study treatment relationship by primary SOC, PT, maximum CTCAE grade
- Most frequent (≥ 5%) AEs, regardless of study treatment relationship by PT, maximum CTCAE grade
- AEs suspected to be related to the study treatment by primary SOC, PT, maximum CTCAE grade
- Most frequent (≥ 5%) AEs, suspected to be related to the study treatment by PT, maximum CTCAE grade
- Serious adverse events (SAE), regardless of study treatment relationship, by primary SOC, PT, maximum CTCAE grade
- SAEs suspected to be related to the study treatment, by primary SOC, PT, maximum CTCAE grade
- Deaths on treatment by primary SOC and PT
- AEs leading to study drug discontinuation, regardless of study treatment relationship, by primary SOC, PT and maximum CTCAE grade
- AEs requiring dose adjustment or study treatment interruption, regardless of study treatment relationship, by primary SOC, PT, maximum CTCAE grade
- AEs requiring additional therapy, regardless of study treatment relationship, by primary SOC, PT, maximum CTCAE grade

In addition, subgroup analyses will be performed for all AEs and AEs suspected to be related to the study treatment.

- Age (< 65 years vs ≥ 65 years, and <75 years vs. ≥ 75 years)
- Race (Asian, Black or African American, White and others)
- Gender (female vs. male)
- Liver metastasis (yes vs. no)
- Prior lines of chemotherapy (1, 2-3 vs. >=4 lines)

AEs of interest will also be summarized. Please see Section 5.1.5 for the grouping details

5.1.5 AEs of special interest

AE of special interest (AESI) consists of adverse events for which there is a specific clinical interest in connection with ribociclib treatment (i.e. where ribociclib 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 AESI 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 AESI can be defined by one or several MedDRA categories. These AESIs are defined in the current version of the ribociclib Safety profiling plan/ ribociclib Case Retrieval Strategy (CRS) document.

The number and percentage of patients will be reported by AESI and by preferred term. Groupings are displayed as below –

- Hepatobiliary toxicity
- QTc prolongation
- Nausea, emesis
- Diarrhea
- Pneumonitis, interstitial lung disease
- Anemia
- Thrombocytopenia
- Haematopoietic cytopenias affecting more than one type of blood cell
- Leukopenia and neutropenia
- Renal Toxicity

Additional AESIs may be reported if there are any updates to the CRS at the time of the analyses.

The number and percentage of patients will be reported by AESI, by MedDRA term, by PT, maximum CTCAE.

These analyses will be repeated in subgroups specified in Section 5.1.4, for the following AESIs:

- Leukopenia and neutropenia
- QT prolongation
- Hepatobiliary toxicity

5.2 Deaths

The number and percentage of patients who died within 30 days of the last dose will be summarized with cause of deaths.

A patient listing of all deaths with recorded principal cause of death (if available) will also be provided. All patients in FAS who died will be included, with deaths that are within 30 days of the last dose noted.

5.3 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.

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 patients 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. Separate listings will also be provided for notable laboratory abnormalities (i.e., grade 3 or 4 laboratory toxicities post start of study treatment for CTC gradable parameters, or abnormal laboratory values for non-CTC gradable parameters). In addition, other lab tests such as those for urinalysis and pregnancy will be listed separately.

Patients' hepatic toxicity will be evaluated for on-treatment liver parameter values. The worst case assessments are used for the analyses. The number and percentage of patients who met the following categories of the criteria will be presented.

- ALT > 3xULN, ALT>5xULN, ALT>8xULN, ALT>10xULN, ALT>20xULN, separately
- AST >3xULN, AST>5xULN, AST>8xULN, AST>10xULN, AST>20xULN, separately
- ALT>3xULN or AST>3xULN, ALT>5xULN or AST>5xULN, ALT>8xULN or AST>8xULN, ALT>10xULN or AST>10xULN, ALT>20xULN or AST>20xULN, separately
- TBILI>ULN, TBILI>2xULN, separately
- ALP>1.5xULN, ALP>2xULN, ALP>3xULN,ALP>5xULN, ALP>8xULN, ALP>10xULN, separately
- (ALT>3xULN or AST>3xULN) and TBILI>ULN, (ALT>5xULN or AST>5xULN) and TBILI>ULN, (ALT>8xULN or AST>8xULN) and TBILI>ULN, (ALT>10xULN or AST>10xULN) and TBILI>ULN, (ALT>20xULN or AST>20xULN) and TBILI>ULN, separately
- (ALT>3xULN or AST>3xULN) and TBILI>2xULN, (ALT>5xULN or AST>5xULN) and TBILI>2x ULN, (ALT>8xULN or AST>8xULN) and TBILI>2x ULN,

(ALT>10xULN or AST>10xULN) and TBILI>2x ULN, (ALT>20xULN or AST>20xULN) and TBILI>2x ULN, separately

- (ALT>3xULN or AST>3xULN) and TBILI>=2xULN and ALP<2xULN

Note: ALT = Alanine Aminotransferase; AST =

5.4 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The parameters expected to be collected include: height (cm), weight (kg), body temperature (°C), heart rate (beats per minute), and systolic and diastolic blood pressure (mmHg).

The number and percentage of patients with clinically notable vital signs changes from baseline will be presented.

Clinically notable vital sign criteria are provided below.

Vital Sign (unit)	**Clinically notable criteria**
Body weight (Kg)	decrease from Baseline of ≥ 10%
	increase from Baseline of ≥ 10%
Systolic blood pressure (mmHg)	≤ 90 and decrease from Baseline of ≥ 20
	≥ 180 and increase from Baseline of ≥ 20
Diastolic blood pressure (mmHg)	≤ 50 and decrease from Baseline of ≥ 15
	≥ 105 and increase from Baseline of ≥ 15
Heart Rate (bpm)	≤ 50 and decrease from Baseline of ≥ 15
	≥ 120 and increase from Baseline of ≥ 15

5.5 ECG

Standard triplicate 12 lead ECG assessments are performed after the patient has been resting for 5-30 min prior to each time point. The average of the triplicate ECGs will be used for each time point. ECG values at each visit per each time point will be summarized descriptively for the SS. Summary parameters are: QTcF, QTcB, HR, PR, QT, and QRS. In addition, the number and percentage of notable abnormal ECGs for the following categories at any time during the study will also be summarized. Shift table base on QTcF, QTcB, and QT categories will be presented from baseline to worst post-baseline values.

ECG parameter (unit)	**Clinically notable criteria**
QT, QTcF, QTcB (ms)	New > 450 ms
	New > 480 ms
	New > 500 ms
	Increase from Baseline > 30 ms

	Increase from Baseline > 60 ms
PR duration (ms)	Increase > 25% from Baseline and to PR duration > 200
QRS duration (ms)	Increase > 25% from Baseline and to QRS duration > 110
Heart Rate (bpm)	≤ 50 and decrease from Baseline of ≥ 25%
	≥ 100 and increase from Baseline of ≥ 25%

5.6 Other safety evaluations

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

Other collected safety parameters, such as cardiac imaging (i.e., LVEF value), will be listed.

6 Interim analyses

Scheduled interim data reviews will occur for the primary endpoint of clinical benefit rate only as required by the Bayesian Hierarchical design. The first interim data review will be performed after the first 30 patients overall (across all groups) have been dosed for at least 16 weeks or discontinued, and then every 13 weeks thereafter. Interim analyses may be performed more frequently dependent on enrollment rate to avoid over-enrollment in any of the disease cohorts. 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 data reviews. In the event of rapid enrollment, interim analyses for publication or other purpose may not be implemented

7 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 Protocol Appendix R Table 4.2.1 and Table 4.2.2. 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 the appendix. The sample sizes shown (minimum of 10 for futility stopping, minimum of 15 for early success and maximum of 30 as a 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 appendix.

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

8 Bayesian Adaptive Design

8.1 Introduction

This document outlines the adaptive design framework to be used for all trials within Novartis's Modular Phase II study to link targeted therapy to patients with pathway activated tumors.

Although the selected tumor types may vary by trial, this document outlines the design and analysis approach based on 8 example tumor cohorts:

- Lung NSC
- Bladder
- Breast
- Colorectal
- GIST
- HNSCC
- Ovarian
- Sarcoma

Other tumor types may be considered for trials if 1) 4 or more patients are enrolled in the cohort, and 2) a reasonable estimate of the clinical benefit rate is available.

The primary endpoint is clinical benefit rate (CBR) in each cohort, with clinical benefit being assessed at 16 weeks. All patients will receive the experimental treatment for that particular trial.

8.1.1 Primary Analysis

We let Y_i be the response indicator for the i^{th} patient, 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 trial. We transform to the logit scale for modeling purposes. Let θ_g be the mean log odds treatment effect, i.e.:

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

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 for that group. Thus, we wish to test the set of hypotheses

$$H_{0g} : \theta_g \leq 0$$

$$H_{1g} : \theta_g > 0$$

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,

$$Pr(\theta_g>0 \mid data) > 0.80$$

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.

8.1.2 Trial Logistics

The trial will enroll all available patients in all cohorts for 2 years unless a cohort cap is reached or a cohort is stopped early, or the trial is stopped early by Novartis. The trial will enroll no more than 30 evaluable patients in each cohort. Interim monitoring will be conducted starting after the first 30 patients are enrolled overall (across all cohorts), and continuing each 13 weeks thereafter till study enrollment closure. After that, one CBR analysis will be done at the end of the study. At each interim data review, response information for the various groups will be evaluated to determine the current $Pr(\theta_g>0 \mid data)$ within each cohort, with sufficiently high/low values used to stop the cohort for success/futility. A minimum of 10 patients will be required in a cohort before it may discontinue enrollment for futility, and a minimum of 15 patients are required before discontinuing a cohort for efficacy. If a cohort stops enrolling early, the remaining cohorts 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 the analysis of the analysis of the study data for the primary CSR.

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

Patients within any cohort which does not reach the minimum 3 patient enrollment will be excluded from the interim and final analyses. As the study continues, early interim data reviews may be based on fewer cohorts than later interim data reviews, as the interim data reviews will include whatever cohorts have satisfied the criteria at the time of the data review.

8.2 Statistical Modeling

We let Y_i be the response indicator for the i^{th} patient, 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 (discussed in more detail in Section 8.5) 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.

8.2.1 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 Section 8.5.

8.3 Evaluation of Trial Success and Futility

Interim monitoring will occur after the first 30 patients are on study for 16 weeks, then every 13 weeks thereafter till study enrollment closure. At each interim data review, 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.

8.3.1 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.

8.3.2 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 patients will need to be evaluated prior to declaring a group to be efficacious.

8.3.3 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 80% probability that the response rate in a group exceeds the historical rate, then the group will be considered a success. Formally:

$$\Pr(\pi_g > R_g) > 0.80.$$

8.4 Simulation

We evaluated type I error and power for each of the 8 possible groups under a variety of possible “truths” indicating various possible true underlying probabilities within each group.

8.4.1 Assumptions

Accrual – Two scenarios for the assumed two-year expected accrual are investigated: 1) 10 patients per group and 2) 5 patients per group. Note that 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.

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

Control Rates – Table 1.1 shows the assumed control clinical benefit rates for each group.

Table 1.1 Assumed control rates

Tumor Type	**Assumed Control Rate (R_g)**
Lung NSC	0.45
Bladder	0.47
Breast	0.50
Colorectal	0.38
GIST	0.50
HNSCC	0.45
Ovarian	0.47
Sarcoma	0.40

Table 1.1 – 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 1.2. Values identical to the control are shown in bold, while values greater than the assumed control rate are italicized.

Table 1.2 Treatment rates

	Null	Alternative	Two	Half
Lung NSC	**0.45**	*0.71*	**0.45**	**0.45**
Bladder	**0.47**	*0.73*	**0.47**	**0.47**
Breast	**0.50**	*0.75*	**0.50**	**0.50**
Colorectal	**0.38**	*0.65*	**0.38**	**0.38**
GIST	**0.50**	*0.75*	**0.50**	*0.75*
HNSCC	**0.45**	*0.71*	**0.45**	*0.71*
Ovarian	**0.47**	*0.73*	*0.73*	*0.73*
Sarcoma	**0.40**	*0.67*	*0.67*	*0.67*

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.

8.4.2 Results

A total of 8 scenarios were simulated (two accrual scenarios and four possible 'truths' for the clinical benefit rate). The probability of group success for each group is provided for each scenarios in the tables below.

Table 2.1 Probability of group success for each cohort assuming expected accrual of 10 patients/cohort

Two-year expected accrual: 10 patients/cohort				
Group(Exp N)	Null	Alternative	Two	Half
Lung NSC	**0.158**	*0.915*	**0.208**	**0.305**
Bladder	**0.131**	*0.918*	**0.232**	**0.322**
Breast	**0.147**	*0.909*	**0.233**	**0.312**
Colorectal	**0.138**	*0.921*	**0.200**	**0.276**
GIST	**0.162**	*0.921*	**0.233**	*0.826*
HNSCC	**0.139**	*0.906*	**0.217**	*0.834*
Ovarian	**0.145**	*0.929*	*0.786*	*0.829*
Sarcoma	**0.135**	*0.939*	*0.758*	*0.852*

Table 2.2 Probability of group success for each cohort assuming expected accrual of 5 patients /cohort

Two-year expected accrual: 5 patients/cohort				
Group	Null	Alternative	Two	Half
Lung NSC	**0.132**	*0.803*	**0.204**	**0.258**
Bladder	**0.140**	*0.830*	**0.196**	**0.265**
Breast	**0.160**	*0.807*	**0.232**	**0.261**
Colorectal	**0.135**	*0.794*	**0.194**	**0.278**
GIST	**0.155**	*0.826*	**0.212**	*0.688*
HNSCC	**0.140**	*0.820*	**0.190**	*0.657*
Ovarian	**0.151**	*0.819*	*0.587*	*0.652*
Sarcoma	**0.139**	*0.799*	*0.579*	*0.667*

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. Italicized entries appear where the treatment is effective, and thus indicate the power of the design.

Generally, type I error is controlled at 0.20 under the null scenario (the borrowing compensates for the multiple interim data reviews) and power is an increasing function of the expected sample size (power is higher in the higher accrual situation across treatment rate scenarios). 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. Note in any particular trial there should be a mix of high and low enrolling cohorts, thus some cohorts may enroll closer to 10 patients while other may only enroll five. This would produce a power value somewhere between the two tables.

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

8.5 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.

9 References

1 **Clopper, C., Pearson, E. S.** The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934; 26: 404-413.

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

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