NOVARTIS

Clinical Development

INC424/ruxolitinib/JAKAVI

CINC424C2301 / NCT02913261

A phase III randomized open-label multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease after allogeneic stem cell transplantation

Statistical Analysis Plan (SAP)

Author: Trial Statistician, [redacted]

Document type: SAP Documentation

Document status: Final

Release date: 25-May-2021

Number of pages: 57

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)

Table of contents

List of abbreviations

AE	Adverse event
AESI	Adverse Event of Special Interest
aGvHD	Acute Graft vs. Host Disease
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
BAT	Best Available Therapy
bid	bis in diem/twice a day
BMI	Body Mass Index
BOR	Best Overall Response
BSA	Body Surface Area
cGvHD	chronic Graft vs. Host Disease
CIBMTR	Center for International Blood and Marrow Transplant Research
CMH	Cochran–Mantel–Haenszel
CMV	Cytomegalovirus
CR	Complete Response
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DAR	Dose Administration Record
DMC	Data Monitoring Committee
eCRF	Electronic Case Report Form
EFS	Event-Free Survival
FACT-BMT	Functional Assessment of Cancer Therapy - Bone Marrow Transplantation
FAS	Full Analysis Set
FFS	Failure-Free Survival
GvHD	Graft vs. Host Disease
HCT	Hematopoietic Cell Transplantation
IRT	Interactive Response Technology that includes Interactive Voice Response System and Interactive Web Response System
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NRM	Non Relapse Mortality
ORR	Overall Response Rate
OS	Overall Survival
PAS	Pharmacokinetic analysis set
PD	Pharmacodynamics
PK	Pharmacokinetics
PPS	Per-Protocol Set
PR	Partial response
PRO	Patient-reported Outcomes
QoL	Quality of Life
RAP	Report and Analysis Process

RPSFT	Rank-Preserving Structure Failure Time
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
WBC	White Blood Cells
SOC	System Organ Class
TFLs	Tables, Figures, Listings
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes all planned analyses for the Clinical Study Report (CSR) of study CINC424C2301, a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease (aGvHD Grade II-IV) after allogeneic stem cell transplantation.

The content of this SAP is based on protocol CINC424C2301 version 02. All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock of the study data.

1.1 Study design

This is a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease after allogeneic stem cell transplantation. Approximately 308 patients will be randomized to one of the following treatment arms in 1:1 ratio:

- Ruxolitinib
- BAT

Randomization will be stratified by the aGvHD grade (Grade II vs. Grade III vs. Grade IV).

Overall response rate (ORR) at Day 28, as assessed by local investigators' review of aGvHD response and using standard criteria [Harris 2016], is the primary endpoint in this study. Durable overall response rate at Day 56 is the key secondary endpoint.

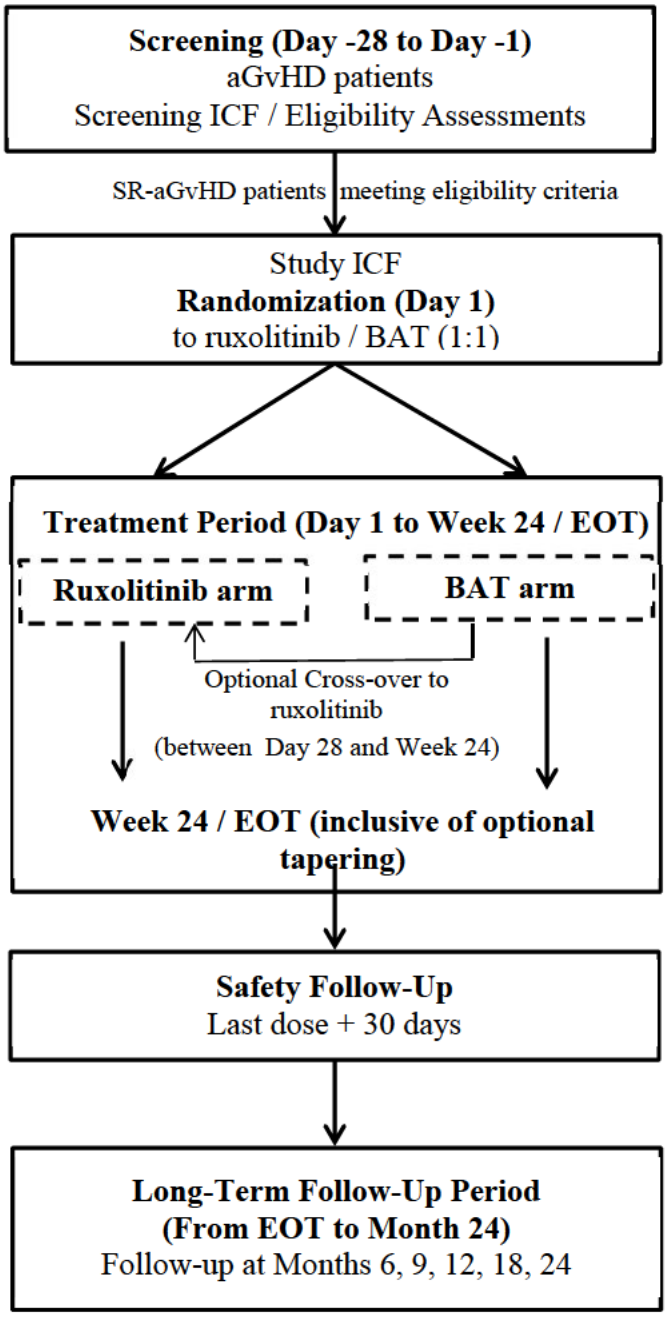
The primary analysis including the analysis on primary and key secondary endpoints was performed with the cut-off date as 25-Jul-2019 after all patients have completed their Day 56 visit or have discontinued study. The primary analysis data were summarized in the primary clinical study report (CSR).

Further analyses on secondary endpoints was performed with the cut-off date as 06-Jan-2020 when all patients have completed 6 months treatment or discontinued from study participation earlier.

The final analysis will occur once all patients have completed the study (up to 24 months from randomization). All available data from all patients up to end of study, inclusive of OS, will be reported in a final CSR.

No formal interim efficacy analysis is planned in this study.

Figure 1-1 Schematic Study Design



1.2 Study objectives and endpoints

Table 1-1 Objectives and related endpoints

Objective	Endpoint	Analysis
Primary		**Refer to Section 2.5**
To compare the efficacy of ruxolitinib vs. Investigator's choice Best Available Therapy (BAT) in patients with Grade II-IV SR-aGvHD assessed by Overall Response Rate (ORR) at Day 28	Overall response rate (ORR) at Day 28 after randomization, defined as the proportion of patients in each arm demonstrating a complete response (CR) or partial response (PR) without requirement for additional systemic therapies for an earlier progression, mixed response or non-response. Scoring of response will be relative to the organ stage at the time of randomization.	CMH analyses, stratified by aGvHD grade
Key secondary endpoint		**Refer to Section 2.6**
To compare the rate of durable ORR at Day 56 between ruxolitinib and BAT	Proportion of all patients in each arm who achieve a complete response (CR) or partial response (PR) at Day 28 and maintain a CR or PR at Day 56.	CMH analyses, stratified by aGvHD grade
Other secondary endpoints		**Refer to Sections 2.7, 2.8, 2.9, 2.10, 2.11**
To estimate ORR at Day 14	Proportion of patients who achieved ORR (CR+PR) at Day 14.	
To assess Duration of response	Duration of response (DOR) is assessed for responders only and is defined as the time from first response until aGvHD progression or the date of additional systemic therapies for aGvHD. Onset of chronic GvHD, or death without prior observation of aGvHD progression are considered as competing risks.	

Objective	Endpoint	Analysis
To assess the cumulative steroid dose until Day 56	Weekly cumulative steroid dose for each patient up to Day 56 or end of treatment will be calculated.	
To assess Overall Survival (OS)	Overall survival, defined as the time from the date of randomization to the date of death due to any cause.	Probability of overall survival at Months 1, 2, 6, 12, 18 & 24 will be estimated from the Kaplan-Meier curves for each arm.
To assess Event-Free Survival (EFS)	Event-free survival, defined as the time from the date of randomization to the date of hematologic disease relapse/progression, graft failure, or death due to any cause.	Probability of EFS at Months 1, 2, 6, 12, 18 & 24 will be estimated from the Kaplan-Meier curves for each arm.
To assess Failure-Free Survival (FFS)	Failure-free survival, defined as the time from the date of randomization to date of hematologic disease relapse/progression, non-relapse mortality, or addition of new systemic aGvHD treatment.	Cumulative incidence of FFS at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering each event as a competing risk for the other two. Onset of chronic GvHD is considered as a competing risk.
To assess Non Relapse Mortality (NRM)	Non-relapse mortality (NRM), defined as the time from date of randomization to date of death not preceded by hematologic disease relapse/progression.	Cumulative incidence of NRM at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering hematologic disease relapse/progression as competing events.
To assess incidence of Malignancy Relapse/Progression (MR)	Malignancy Relapse/Progression (MR) (refer to protocol Appendix 4), defined as the time from date of randomization to hematologic malignancy relapse/progression. Calculated for patients with underlying hematologic malignant disease.	Cumulative incidence of MR at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering deaths not preceded by hematologic malignancy relapse/progression as competing events.
To measure the incidence of cGvHD	cGvHD, defined as the diagnosis of any cGvHD including mild, moderate, severe.	Cumulative incidence of cGvHD at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering hematologic disease relapse/progression and death without prior cGvHD as competing events.
To estimate the rate of Best Overall Response (BOR)	Proportion of patients who achieve OR (CR+PR) at any time point up to and including Day 28 and	

Objective	Endpoint	Analysis
	before the start of additional systemic therapy for aGvHD.	
To assess Pharmacokinetics (PK) of ruxolitinib in SR-aGvHD patients	Pharmacokinetic parameters of ruxolitinib after a single dose and at steady state. Cmax, AUClast, AUCinf, Ctrough, Racc and AUCtau. Other PK parameters are CL/F, Vz/F, Tmax and T1/2.	
To assess exposure-response relationship of ruxolitinib in SR-aGvHD	Pharmacokinetics (exposure) and efficacy (ORR, OS, or other relevant endpoints) relationship. Pharmacokinetics (exposure) and safety (AEs) relationship.	
To evaluate changes in Patient Reported Outcomes (PROs)	Change in FACT-BMT from baseline to each visit where measured. Change in EQ-5D-5L from baseline to each visit where measured.	
To evaluate the safety of ruxolitinib and Best Available Therapy	Safety and tolerability including myelosuppression, infections, and bleeding will be assessed by monitoring the frequency, duration and severity of Adverse Events including occurrence of any second primary malignancies, infections, by performing physical exams, and evaluating changes in vital signs from baseline, routine serum chemistry, hematology results and coagulation profile.	
[redacted]		[redacted]
[redacted]	[redacted]	

Objective	Endpoint	Analysis
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
	[redacted]	
	[redacted]	

2 Statistical methods

2.1 Data analysis general information

The final analysis will be performed by Novartis. SAS version 9.4 or later and/or R version 3.0.2 or later will be used to perform all data analyses and to generate tables, figures and listings.

Data included in the analysis

The analysis cutoff date for the final analysis of study data will be established when all patients have completed the study.All statistical analyses will be performed using all data collected in the database up to the data cutoff date. All data with an assessment date or event start date (e.g. vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'ongoing'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these events, the end date will not be imputed and therefore will not appear in the listings.

General analysis conventions

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

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

Quantitative data (e.g., age, body weight, etc.) will be summarized by appropriate

descriptive statistics (i.e. mean, standard deviation, median, minimum, and maximum) by treatment group.

2.1.1 General definitions

Investigational drug and study treatment

Investigational drug, will refer to the ruxolitinib only. Whereas, ***study treatment*** will refer to ruxolitinib and BAT.

Randomized treatment will refer to the study treatment received during the randomized treatment period. Up to Day 28 visit, more than one BAT regimen may be initiated as study treatment. These regimens reported on the Dosage Administration Record (DAR) eCRF are considered randomized treatment. ***Crossover treatment*** will refer to the study treatment received during the crossover treatment period.

Date of first administration of randomized treatment

The date of first administration of randomized treatment is derived as the first date when a nonzero dose of randomized treatment was administered as per the DAR eCRF. The date of first administration of randomized treatment will also be referred as ***start of randomized treatment***.

Date of first administration of crossover treatment

The date of first administration of crossover treatment is derived as the first date when a nonzero dose of crossover treatment was administered as per the DAR eCRF. The date of first administration of crossover treatment (ruxolitinib) will also be referred as ***start of crossover treatment***.

Date of last administration of randomized treatment

The date of last administration of randomized treatment is defined as the last date when a nonzero dose of randomized treatment was administered as per DAR eCRF.

Date of last administration of crossover treatment

The date of last administration of crossover treatment is defined as the last date when a nonzero dose of crossover treatment (ruxolitinib) was administered as per DAR eCRF.

Study day

The study day, describes the day of the event or assessment date, relative to the reference start date.

The study day is defined as:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

The reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, PK, etc.), [redacted] and PRO is the start of randomized treatment.

The reference start date for all other, non-safety assessments (i.e., aGvHD assessment, survival, aGvHD progression, aGvHD response, underlying hematologic disease relapse/progression, etc.) is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Crossover study day

The crossover study day, describes the day of the event or assessment date, relative to the start of crossover treatment.

Crossover study day = date of event – start of crossover treatment + 1, if event is on or after the start of crossover treatment

Crossover study day = date of event – start of crossover treatment, if event precedes the start of crossover treatment

The crossover study day will be displayed in the data listings if an event starts on or after the start of crossover treatment.

Time unit

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

Baseline

Baseline for safety endpoints (except adverse events), [redacted] and patient reported outcomes is defined as the last assessment prior to or on the treatment start date.

For evaluations after cross over, the baseline is defined as the last assessment prior to or on the start date of crossover treatment.

Baseline for efficacy endpoints is defined as the last assessment or procedure conducted prior to or on the date of randomization date + 3 days, but no later than the treatment start date.

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

On-treatment assessment/event and observation periods

For adverse event reporting the overall observation period will be divided into four mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent at screening to the day before first administration of study treatment
2. ***on-randomized-treatment period***: from date of first administration of randomized treatment to 30 days after date of last actual administration of randomized treatment (including start and stop date), or end of randomized treatment per End of Randomized Treatment Disposition eCRF, whichever is later; for those patients who cross over from BAT to ruxolitinib, the period is from date of first administration of randomized treatment to earlier of (i) 30 days after date of last actual administration of randomized treatment, or end of randomized treatment per End of Randomized Treatment Disposition eCRF, whichever is later, (ii) the day before the date of first administration of crossover treatment. Up to Day 28 visit, more than one BAT regimen may be initiated as study treatment. In this case, the last actual administration of randomized treatment refers to the last actual administration of the last BAT regimen reported on the DAR eCRF.
3. ***on-crossover-treatment period***: from date of first administration of crossover treatment to 30 days after date of last administration of crossover treatment (including start and

stop date), or end of crossover treatment per End of Crossover Treatment Disposition eCRF, whichever is later.

4. ***post-treatment period***: starting at Day 31 after last administration of study treatment, or the day after end of study treatment per end of treatment disposition eCRFs, whichever is later.

If dates are incomplete in a way that clear assignment to pre-, on-, post-treatment period cannot be made, then the respective data will be assigned to the on-randomized-treatment period.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). The data on-randomized-treatment period and on-crossover-treatment period will be summarized separately. In addition, a separate summary for death including on-treatment and post-treatment deaths will be provided. In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period (***treatment-emergent*** AEs).

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

Windows for multiple assessments

In order to summarize data collected over time (including unscheduled visits), the assessments will be time slotted. The following general rule will be applied in creating the assessment windows: If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If two assessments within a time window are equidistant from the target date, then the later of the two assessments will be used. If multiple assessments on the same date then the worst case will be used. Data from all assessments (scheduled and unscheduled), including multiple assessments, will be listed.

The following time windows are defined for descriptive summary on aGvHD assessment, [redacted] PROs and safety (Table 2-1) by visit. The end of treatment assessment will be mapped into the time points if collected within 7 days of the last dose intake.

Table 2-1 Time windows for aGvHD assessment, [redacted] PROs and safety assessment (lab, vital sign, etc.)

Time Window	Planned Visit Timing	Time Window Definition
On treatment		
Baseline[1]	Study Day 1[1]	≤ Study Days 1+3, no later than treatment start date
Baseline[2]	On or before Study Day 1[2]	≤ Study Day 1
Week 1	Study Day 7	Study Days 4 – 10
Week 2	Study Day 14	Study Days 11 – 17
Week 3	Study Day 21	Study Days 18 – 24
Week 4	Study Day 28	Study Days 25 – 31
Week 5	Study Day 35	Study Days 32 – 38
Week 6	Study Day 42	Study Days 39 – 45

Week 7	Study Day 49	Study Days 46 – 52
Week 8	Study Day 56	Study Days 53 – 59
Week 12	Study Day 84	Study Days 71 – 98
Week 16	Study Day 112	Study Days 99 – 126
Week 20	Study Day 140	Study Days 127 – 154
Week 24	Study Day 168	Study Days 155 – 182
Safety follow-up	30 days after last dose	Last dose date + 30

Baseline[1] for aGvHD assessment (efficacy);
Baseline[2] for PROs, safety assessments [redacted] (safety);
Study Day 1[1] = randomization date;
Study Day 1[2] = start date of randomized treatment;
Crossover Study Day 1 = start date of crossover treatment
EOT (randomized treatment or crossover treatment) assessments are mapped to the time points.
Safety follow-up is a separate time point for PROs and safety assessments.

Last contact date

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

Table 2-2 Last contact date data sources

Source data	Conditions
Date of Randomization	No Condition
Last date patient was known to be alive from Survival Follow-up page	Patient status is reported to be alive, lost to follow-up or unknown.
Start/End dates from further aGvHD therapy	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
End of treatment date from end of treatment page	No condition.
- aGvHD assessment date - any specific efficacy assessment date if available (e.g., cGvHD assessment, graft failure assessment, hematologic disease relapse/progression assessment)	Evaluation is marked as 'done'.
Laboratory/PK collection dates/[redacted]	Sample collection marked as 'done'.
Vital signs date	At least one non-missing parameter value
Start/End dates of AE	Non-missing verbatim term

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date . The cut-off date will not be used for last contact date, unless the patient was seen or contacted on that date. No date post cut-off date will be used. Completely 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. Partial date imputation is allowed for event (death)/censoring coming from 'Survival information' eCRF.

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

2.2 Analysis sets

Full Analysis Set

The Full Analysis Set (FAS) comprises all patients to whom study treatment has been assigned by randomization. According to the intent to treat principle, patients will be analyzed according to the treatment and strata, they have been assigned to during the randomization procedure.

Per-Protocol Set

The Per-Protocol Set (PPS) consists of a subset of the patients in the FAS who are compliant with requirements of the clinical study protocol.

The following list of protocol deviations will lead to exclusion of the patient from the Per-Protocol Set:

- Not steroid refractory aGvHD (PDID: INCL09 per SSD v22.0)
- More than one prior systemic therapy for the treatment of aGvHD other than corticosteroids +/- CNI (prophylaxis or treatment) (PDID: EXCL01 per SSD v22.0)
- Missing or incorrect aGvHD grade at randomization (PDID: OTH02, INCL07 per SSD v22.0)
- Taking any prohibited medication as specified in this protocol after start of study treatment and before end of study treatment (PDID: COMD01, COMD02 per SSD v22.0)
- Study treatment received different from treatment arm assigned by randomization (PDID: TRT02 per SSD v22.0)

Safety Set

The Safety Set includes all patients who received at least one dose of study treatment. Patients will be analyzed according to the study treatment received, where treatment received is defined as the randomized treatment if the patient took at least one dose of that treatment or the first treatment received if the randomized treatment was never received during the randomized treatment period.

Crossover Analysis Set

The Crossover Analysis Set (CAS) comprises all patients randomized to and received BAT who then crossed over and received at least one dose of ruxolitinib. This analysis set will be used for all analyses for crossover patients.

Pharmacokinetic Analysis Set

The Pharmacokinetic Analysis Set (PAS) includes all patients who provide at least one evaluable PK concentration. For a concentration to be evaluable, patients are required to:

- Take a dose of ruxolitinib prior to sampling.
- For pre-dose samples, do not vomit within 2 hours after the dosing of ruxolitinib prior to sampling; for post-dose samples, do not vomit within 2 hours after the dosing of ruxolitinib.

The PAS will be used for NCA analysis for patients where extensive PK sampling is obtained [redacted]

Patient Classification:

Patients may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific patient classification rules defined in Table 2-3.

Table 2-3 Patient classification based on protocol deviations and non-PD criteria

Analysis set	Protocol deviations leading to exclusion	Non protocol deviation leading to exclusion
FAS	INCL03	Not applicable
Safety Set	INCL03	No dose of study treatment
Per-Protocol Set	INCL03, INCL07, INCL09, EXCL01, TRT02, , OTH02, During randomized treatment period: COMD01, COMD02	No dose of study treatment
Crossover Analysis Set	INCL03	No dose of ruxolitinib
PK Analysis Set	INCL03	No dose of ruxolitinib, No evaluable PK concentration

Note: Based on CINC424C2301_Study Specification Document version 22.0.
INCL03 - Written Study informed consent /assent not obtained.

Withdrawal of Informed Consent

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

2.2.1 Subgroup of interest

Efficacy

The primary efficacy endpoint will be summarized by the following subgroups to examine the homogeneity of treatment effect provided that the primary efficacy analysis based on the FAS is statistically significant :

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race
- Region (Europe, Australia and Canada), Japan, Asia excluding Japan)
- Acute GvHD grade (Grade II, III, IV)
- Source of grafts (related, not related)
- Criteria for SR-aGvHD (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper)
- Prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/- other systemic aGvHD treatment,)
- Conditioning regimen type (myeloablative, non-myeloablative, reduced intensity)
- Stem cell type (bone marrow, peripheral blood, single cord blood)
- Donor HLA status (match, mismatch)
  - Note: HLA status Match score means that 1:1 match, e.g. 10/10, 8/8, 6/6, 12/12. Mismatch will be all the other scores.
- Donor gender match (F/F, M/M, F/M, M/F)
- Donor CMV status (negative, positive)
- Donor source/HLA match status (related and matched, related and mismatched, not related and matched, not related and mismatched)
- aGvHD organ involvement at randomization (skin, liver, upper GI, lower GI)

No formal statistical test of hypotheses will be performed for the subgroups, only point estimate of the treatment effect and 95% confidence intervals will be provided (see Section 2.5 for further analysis details). The objective of the efficacy subgroup analysis is to demonstrate homogeneity of treatment effect in the above subgroups.

Safety

Key safety analyses will be repeated on the Safety Set in the following subgroups:

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race
- Region (Europe, Australia and Canada), Asia excluding Japan, Japan)
- aGvHD organ involvement at randomization (skin, liver, upper GI, lower GI)

The objective for carrying out these subgroup analyses is to identify potential safety issues that may be limited to or more commonly observed in a subgroup of patients. The following summaries will be presented by subgroup:

- AEs, irrespective of causality, by primary system organ class and preferred term
- AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- Serious AEs, irrespective of causality, by primary system organ class and preferred term
- Serious AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- On-treatment deaths, by primary system organ class and preferred term

Adolescent patients

In order to be able to make a separate risk/benefit assessment for the adolescent patients, besides the above planned subgroup analyses for this age group (12-<18 years), data of demographics and exposure will be presented.

Japanese patients

Subgroup analyses will also be performed for the patients treated in Japan. No selection will be done on the basis of ethnicity, the purpose being to evaluate the population of patients living in Japan, not a specific ethnic set of patients. The analysis will be done the same way as in region subgroup specified above.

2.3 Patient disposition, demographics and other baseline characteristics

The Full Analysis Set (FAS) will be used for all baseline and demographic summaries and listings unless otherwise specified. Summaries will be reported by treatment arm and for all patients and listings will be reported by treatment arm to assess baseline comparability. No inferential statistics will be provided.

Basic demographic and background data

All demographic and baseline disease characteristics data will be summarized and listed by treatment arm. Categorical data (e.g. gender, age groups: 12- <18 vs. 18-65 vs. >65 years, race, ethnicity) will be summarized by frequency counts and percentages; the number and percentage of patients with missing data will be provided. Continuous data (e.g. age, weight, height, body mass index) will be summarized by descriptive statistics (N, mean, median, standard deviation, Q1, Q3, minimum and maximum).

Baseline stratification factors

The number (%) of patients in each stratum (Grades II, III, IV) based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS. Discordances between the stratum recorded in IRT at the time of randomization and the actual stratum recorded in the clinical database through the data collected on eCRF will be cross-tabulated and listed.

Diagnosis and extent of disease

Summary statistics will be tabulated for diagnosis and extend of disease in underlying disease, stem cell transplant and acute GvHD.

For underlying disease, the analysis will include the following: primary diagnosis category and subcategory, details of primary diagnosis, time since diagnosis of underlying disease, CIBMTR risk assessment.

For transplant related disease history, the analysis will include the following: conditioning regimen type, total HCT-specific comorbidity index score, time since transplant, time from diagnosis of underlying disease to transplant, stem cell type, cytomegalovirus status, donor information including age, gender, HLA typing method, HLA match score, source of grafts (related/unrelated), CMV status, T-cell depleted (Y/N), total nucleated cell dose.

For aGvHD disease history, the analysis will include the following: time since diagnosis of aGvHD grade ≥2, aGvHD grade when diagnosis of grade ≥2, steroid refractory aGvHD criteria met (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper), prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/-other systemic aGvHD treatment,), time from diagnosis of aGvHD grade ≥2 to steroid refractory, time since steroid refractory aGvHD, aGvHD grade at randomization, aGvHD organ involvement, steroid dose at randomization.

Medical history

Medical history and ongoing conditions, including underlying disease conditions and symptoms entered on eCRF will be summarized and listed by treatment arm. The summaries will be presented by primary system organ class (SOC), preferred term (PT) and treatment arm. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

Other

Intended BAT strategy in case of SR aGvHD prior to randomization will be summarized by treatment arm.

All data collected at baseline including child bearing potential will be listed.

2.3.1 Patient disposition

Enrollment by country and center will be summarized for all screened patients and also by treatment arm. The number (%) of randomized patients will be presented overall and by treatment group. The number (%) of screened and not-randomized patients and the reasons for screening failure will also be displayed. The number (%) of patients in the FAS who are still on treatment, who discontinued the treatment phases (randomized treatment, crossover treatment) as well as the reason for discontinuation, and the survival follow-up will be presented overall and by treatment group.

Protocol deviations

The number (%) of patients in the FAS with any protocol deviation will be tabulated by deviation category (as specified in the study Data Handling Plan) overall and by treatment group for the FAS. All protocol deviations will be listed.

Analysis sets

The number (%) of patients in each analysis set (defined in Section 2.2) will be summarized by treatment group and stratum.

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

2.4.1 Study treatment / compliance

Duration of exposure in days to ruxolitinib and BAT will be summarized by means of descriptive statistics. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. Duration of exposure to each BAT regimen (anti-thymocyte globulin (ATG), extracorporeal photopheresis (ECP), mesenchymal stromal cells (MSC), low-dose methotrexate (MTX), mycophenolate mofetil (MMF), mTOR inhibitors (everolimus or sirolimus), etanercept, or infliximab) will be summarized using the same approach.

Actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized for the ruxolitinib arm. Patients randomized to Investigator's choice of BAT will receive various different categories of therapy.

The number (%) of patients who have dose changes or interruptions, and the reasons, will be summarized for ruxolitinib group.

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

The Safety Set and Crossover Analysis Set will be used for all summaries on randomized treatment and crossover treatment, respectively. The Safety Set will be used for listings of study treatment.

Duration of exposure to study treatment

Duration of exposure to randomized treatment (days) = (last date of exposure to randomized treatment) – (date of first administration of randomized treatment) + 1.

Duration of exposure to crossover treatment (days) = (last date of exposure to crossover treatment) – (date of first administration of crossover treatment) + 1.

The last date of exposure to study treatment is the latest of the last dates of exposure to study treatment (see Table 2-4).

Table 2-4 Definition of last date of exposure of study treatment

Scenario	Definition of last date of exposure of study treatment	Example
Scenario 1: Study treatment with a periodical administration	The planned end date of the last period in which the last non-zero dose of the study treatment was last administered.	Example 1: In a once-a-week administration, the last date of exposure is the date of last administration + 6 days.

	Note : If the patient changed to another BAT, died or was lost to follow-up before the derived last date, the last date of exposure to study treatment is the date prior to the start of the next BAT, date of death or the date of last contact, respectively. If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	Example 2: In a twice-a-week administration, the last date of exposure is the date of last administration + 3 days.
Scenario 2: Study treatment with daily/IV administration	Date of last administration of a non - zero dose of the study treatment.	Example 3: A patient had a permanent discontinuation of the study treatment 06Jan2017 after being put on a temporary interruption since 01Jan2017. In this case the last date of exposure is- 31Dec2016.
Scenario 3: Study treatment as an antibody	Date of last administration of a non - zero dose of the study drug + number of days antibody persists in vivo – 1 day. Note : If the patient changed to another BAT, died or was lost to follow-up before the derived last date, the last date of exposure to study treatment is the date prior to the start of the next BAT, date of death or the date of last contact, respectively. If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	Example 4: For a study treatment which antibody persists in vivo for 28 days, the last date of exposure is the date of last administration + 28 days – 1 day.

Summary of duration of exposure of study treatment in days will include categorical summaries (based on clinically meaningful time intervals) and continuous summaries (i.e. mean, standard deviation etc.) using appropriate units of time.

Duration of exposure in patient-years

The duration of exposure in patient-years is a total of the duration of exposure in years from all the patients in a treatment group. It will be calculated for randomized treatment (by treatment group) and crossover treatment, respectively.

Duration of treatment period

Duration of randomized treatment period (days) = end date of on-randomized-treatment period – date of first administration of randomized treatment + 1

Duration of crossover treatment period (days) = end date of on-crossover-treatment period – date of first administration of crossover treatment + 1

The on-randomized-treatment period and on-crossover-treatment period are defined in Section 2.1.1.

The duration of randomized treatment period in days for ruxolitinib and BAT will be summarized by means of descriptive statistics. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The duration of crossover treatment period will be summarized similarly. As patients randomized to BAT may crossover between Day 28 and Week 24, the time from the randomization to the first dose of crossover treatment will also be summarized.

Duration of treatment period in patient-years

The duration of treatment period (randomized and crossover) in patient-years is a total of the duration of treatment period in years from all the patients in a treatment group. It will be calculated for randomized treatment (ruxolitinib vs. BAT) and crossover treatment separately.

Cumulative dose

The **planned cumulative dose** for ruxolitinib refers to the total planned dose as per the protocol (10 mg bid) up to the last dose date.

The **actual cumulative dose of randomized ruxolitinib** refers to the total actual dose of randomized ruxolitinib as documented in the DAR eCRF.

The **actual cumulative dose of crossover ruxolitinib** refers to the total actual dose administered, over the duration for which the patient is on the crossover ruxolitinib as documented in the DAR eCRF.

For patients who did not take any drug the cumulative dose is by definition equal to zero.

Dose intensity and relative dose intensity of ruxolitinib

Dose intensity (DI) **of randomized ruxolitinib** for patients with non-zero duration of exposure is defined as follows:

DI (mg / day) = Actual cumulative dose (mg) of randomized ruxolitinib / Duration of exposure to randomized ruxolitinib (days).

Dose intensity (DI) **of crossover ruxolitinib** for patients with non-zero duration of exposure is defined as follows:

DI (mg / day) = Actual cumulative dose (mg) of crossover ruxolitinib / Duration of exposure to crossover ruxolitinib (days).

For patients who did not take any drug the DI is by definition equal to zero.

Planned dose intensity (PDI) is defined as follows:

PDI (mg / day) = Planned Cumulative dose (mg) / Duration of exposure (days).

The protocol planned starting dose for ruxolitinib is 10 mg BID.

Relative dose intensity (RDI) is defined as follows:

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

DI and RDI will be summarized for randomized and crossover ruxolitinib treatment, separately.

The actual cumulative dose, DI and RDI up to Day 28 visit, Day 56 visit and last date of exposure to study treatment (randomized or crossover) will be summarized.

The number (%) of patients at total daily dose 5 mg, 10 mg, 15 mg and 20 mg will be summarized at Day 28 visit, Day 56 visit and last date of exposure to study treatment.

Dose reductions, interruptions or permanent discontinuations

The number of patients who have dose changes, or interruptions, and the reasons, will be summarized for ruxolitinib (randomized and crossover treatment separately). The number of patients who have dose permanent discontinuations and the reasons, will be summarized by treatment group.

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

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

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

2.4.2 Prior, concomitant and post therapies

2.4.2.1 Prior aGvHD treatment

The number and percentage of patients who received any prior aGvHD treatment (medications and procedures) will be summarized by lowest ATC class, preferred term and treatment arm.

Listings will be produced for prior aGvHD treatment.

The above analyses will be performed using the FAS.

2.4.2.2 Prior prophylaxis

The number and percentage of patients who received any prophylaxis prior to randomization will be summarized by lowest ATC class, preferred term and treatment arm using FAS.

The number and percentage of patients who received prior aGvHD therapy that started before aGvHD diagnosis date and continued after diagnosis will be summarized by lowest ATC class, preferred term and treatment arm using FAS.

Listings will be generated for prophylaxis.

2.4.2.3 Systemic corticosteroid

Doses of methylprednisolone will be converted to prednisone equivalents by multiplying the methylprednisolone dose by 1.25.

The weight-adjusted doses will be calculated by dividng the baseline weight (last measurement on or before the randomization).

The duration of exposure and the peak dose will be summarized for pre-randomizatition period, between screening and randomization.

The duration of exposure will be summarized for on-randomized-treatment period and on-crossover-treatment period separately. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The actual cumulative dose, dose intensity and relative dose intensity (relative to the starting dose of corticosteroids) will be summarized up to Day 28 visit, Day 56 visit and end of on-treatment period. These analyses will be based on Safety Set. The data will be reported on the DAR eCRF through the end of randomized treatment per EOT Disposition eCRF, and on the Prior and Concomitant Medications eCRF after the end of randomized treatment. The list of systemic corticosteroid will be documented in Data Handling Plan.

2.4.2.4 Calcineurin inhibitors (CNIs) during study treatment

The duration of exposure will be summarized for CNIs (cyclosporine or tacrolimus) during on-randomized-treatment period and on-crossover-treatment period, respectively. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. These analyses will be based on Safety Set. The data will be reported on the DAR eCRF through the end of randomized treatment per EOT Disposition eCRF, and on the Prior and Concomitant Medications eCRF after the end of randomized treatment. The list of CNIs will be documented in Data Handling Plan.

2.4.2.5 Additional systemic aGvHD therapy

New additional systemic aGvHD therapy (medications and procedures) since start of study treatment will be listed and summarized by lowest ATC class, preferred term, overall and by treatment group by means of frequency counts and percentages using FAS.

2.4.2.6 Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) other than the study treatment administered to a patient coinciding with the study treatment period. Concomitant therapy include medications (other than study drugs) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system and summarized by lowest ATC class and preferred term using frequency counts and percentages. Surgical and medical procedures will be coded using

MedDRA and summarized by SOC and preferred term. The summaries for randomized treatment phase using Safety Set will include:

- Medications starting on or after the start of randomized treatment but no later than end of on-randomized-treatment period; and
- Medications starting prior to start of randomized treatment and continuing after the start of randomized treatment.

The summaries for crossover treatment phase using Crossover Analysis Set will include:

- Medications starting on or after the start of crossover treatment but no later than end of on-crossover-treatment period; and
- Medications starting prior to start of crossover treatment and continuing after the start of crossover treatment.

All concomitant therapies will be listed using Safety Set. Any concomitant therapies starting and ending prior to the start of randomized treatment or starting beyond end of on-randomized-treatment period if not crossed over, or starting beyond end of on-crossover-treatment period if crossed over, will be flagged in the listing.

The prohibited concomitant medications will be summarized by lowest ATC class and preferred term up to the end of on-randomized-treatment and on-crossover-treatment periods, respectively. The list of prohibited medications will be provided and updated regularly by clinical team according to the clinical database review. The topical medications are excluded from the list.

In addition, a subset of concomitant medications i.e. transfusions (red blood cells and platelets) will be grouped and summarized by treatment group.

The aGvHD prophylaxis on or after treatment start will be summarized by lowest ATC class, preferred term and treatment arm using FAS.

2.5 Analysis of the primary objective

The primary objective of the study is to compare the overall response rate (ORR) at Day 28 between the ruxolitinib arm and BAT arm in steroid refractory aGvHD patients.

2.5.1 Primary endpoint

The primary endpoint was reported in the primary analysis, and will not be repeated in this final analysis, except for the subgroup analyses.

ORR at Day 28 is defined as the proportion of patients with complete response (CR) or partial response (PR) at Day 28 according to standard criteria [Harris 2016]. Note that response is relative to the assessment of aGvHD at randomization.

- **Complete response** is defined as a score of 0 for the aGvHD grading in all evaluable organs that indicates complete resolution of all signs and symptoms of aGvHD in all evaluable organs without administration of additional systemic therapies for any earlier progression, mixed response or non-response of aGvHD.
- **Partial response** is defined as improvement of 1 stage in 1 or more organs involved with aGvHD signs or symptoms without progression in other organs or sites without

administration of additional systemic therapies for an earlier progression, mixed response or non-response of aGvHD.

- **Lack of response** is defined as no response, mixed response, or progression.
  - **No response** is defined as absence of improvement in any organ involved by aGvHD, without worsening in any involved organ.
  - **Mixed response** is defined as improvement of at least 1 stage in the severity of aGvHD in at least one organ accompanied by progression in another organ or development of signs or symptoms of aGvHD in a new organ.
  - **Progression** is defined as worsening in 1 or more organs by 1 or more stages without improvement in any involved organ

A flare in aGvHD is defined as any increase in signs or symptoms of aGvHD after an initial response (CR or PR). However, a flare may not lead to progression or additional systemic therapy. Only flares in GvHD that require new additional systemic therapy, will be considered aGvHD flare failure. Patients who fail corticosteroid taper fulfilling either one of the following criteria should initiate additional systemic therapy:

- Requirement for an increase in the corticosteroid dose to methylprednisolone ≥2 mg/kg/day (or equivalent prednisone dose ≥2.5 mg/kg/day), OR
- Failure to taper the methylprednisolone dose to <0.5 mg/kg/day (or equivalent prednisone dose <0.6 mg/kg/day) for a minimum 7 days.

A patient will not be considered a responder at Day 28 if any of the following events occurs:

- Missing aGvHD assessment at baseline or Day 28
- No CR or PR at Day 28
- Additional systemic therapy for aGvHD prior to Day 28

Compared to the protocol version 00, 'discontinuation from randomized treatment' has been dropped as reason. Since some BATs have fixed duration which is less than 8 weeks, to avoid bias, discontinuation from randomized treatment will not be used to disqualify the response or durable response. Also duration of response will not be censored based on treatment discontinuation.

Acute GvHD will be assessed according to standard criteria [Harris 2016], as described in protocol Appendix 1. Grade will be calculated based on the staging of the organs and recorded on the eCRF by the Investigator. The Investigator reported grade will be used for randomization and reported grade and response for all analyses. Grade and response will be calculated by the sponsor for the purposes of data review only and sensitivity analysis.

2.5.2 Statistical hypothesis, model, and method of analysis

The primary efficacy analysis is the comparison of ORR at Day 28 between the two treatment arms. The following statistical hypotheses will be tested to address the primary efficacy objective:

H_0: $ORR_{rux} \leq ORR_{BAT}$ vs. H_1: $ORR_{rux} > ORR_{BAT}$

where ORR_{rux} and ORR_{BAT} are the overall response rates at Day 28 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the

randomization stratification factor (i.e., aGvHD Grade II vs III vs. IV), will be used to compare ORR between the two treatment groups, at the one-sided 2.5% level of significance.

The primary efficacy variable, ORR at Day 28, will be analyzed at the time when all patients have completed their Day 56 visit or discontinued earlier. The primary analysis will be performed on FAS according to ITT principle. ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. One-sided p-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

2.5.3 Handling of missing values/censoring/discontinuations

Patients with missing assessments that prevent the evaluation of the primary endpoint will be considered non-responders on that treatment arm. This includes missing aGvHD response assessments at baseline and Days 28, 56.

The following analysis windows (also in Table 2-1) will be applied to the target day for assessments on overall response, where target day for Week X is X*7.

Baseline assessment is the last aGvHD assessment prior to or on the date of randomization (Day 1). A tolerance of up to 3 days from randomization will be considered in the analysis, but the baseline assessment will not be later than the date of treatment start.

Weeks 1, 2, 3, 4, 5, 6, 7, 8: -3 days/+3 days

Weeks 12 to 24: -13 days/+14 days

The analysis windows for assessments after crossover is similar, except that the baseline is the last aGvHD assessment prior to or on Crossover Day 1 (date of first administration of crossover treatment).

No data imputation will be applied.

2.5.4 Supportive analyses for ORR at Day 28

For each of the subgroups, the following analyses will be performed:

- Proportion of patients with ORR using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]
- Odds ratio with 95% CI using a logistic regression model with treatment and stratification factors as covariate

Efficacy analyses in subgroups will be purely exploratory and are intended to explore the consistency of treatment effect. Forest plot (n, odds ratio, 95% CI) will be produced to graphically depict the treatment effect estimates in different subgroups. No inferential statistics (p-values) will be produced for the subgroups.

2.5.5 ORR at Crossover Day 28

ORR at Crossover Day 28 is defined as the proportion of crossover patients with complete response (CR) or partial response (PR) at Crossover Day 28 according to standard criteria [Harris 2016]. ORR will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] based on the CAS using local investigators' review of aGvHD assessment data. Note that response is relative to the last assessment of aGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a responder at Crossover Day 28 if any of the following events occurs:

- Missing aGvHD assessment at Crossover baseline or Crossover Day 28
- No CR or PR at Crossover Day 28
- Additional systemic therapy for aGvHD prior to Crossover Day 28

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Crossover baseline to Crossover Day 28.

2.6 Analysis of the key secondary objective

The key secondary objective of the study is to determine whether treatment with ruxolitinib has better durable ORR at Day 56 compared with BAT.

2.6.1 Key secondary endpoint

The key secondary endpoint was reported in the primary analysis, and will not be repeated in this final analysis.

Durable ORR at Day 56 is defined as the proportion of all patients in each arm who achieve a complete response (CR) or partial response (PR) at Day 28 and maintain a CR or PR at Day 56. If a patient is a CR at Day 28 and a PR at Day 56, he/she will be considered as a durable responder. A patient will not be considered a durable responder at Day 56 if any of the following events occurs:

- Not a responder at Day 28
- Missing aGvHD assessment at Day 56
- No CR or PR at Day 56.
- Additional systemic therapy for aGvHD prior to Day 56

Durable ORR will be calculated based on the FAS using local investigators review of aGvHD assessment data.

The patients randomized to BAT who meet cross-over criteria and cross-over to ruxolitinib are considered to have the additional systemic therarpy for aGvHD, and will not be considered as a responder afterwards.

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Day 28 to Day 56 value for the patients who achieved PRs at Day 28.

2.6.2 Statistical hypothesis, model, and method of analysis

The following statistical hypotheses will be tested to address the key secondary efficacy objective:

H_0: $DORR_{rux} \leq DORR_{BAT}$ vs. H_1: $DORR_{rux} > DORR_{BAT}$

where $DORR_{rux}$ and $DORR_{BAT}$ are the durable overall response rates at Day 56 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., aGvHD Grade II vs III vs. IV), will be used to compare the durable ORR between the two treatment groups, at the one-sided 2.5% level of significance, if the primary endpoint is significant.

The key secondary efficacy variable, durable ORR at Day 56, will be analyzed at the time when all patients have completed their Day 56 visit or discontinued earlier. The analysis will be based on FAS according to ITT principle. Durable ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. P-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

The durable ORR at Day 56 will be tested hierarchically. That is, if the ORR at Day 28 is statistically significant, the durable ORR at Day 56 will be tested. If the ORR at Day 28 is not statistically significant, the durable ORR at Day 56 will not be tested.

[redacted]

2.6.3 Handling of missing values/censoring/discontinuations

Refer to Section 2.5.3

2.6.4 Durable ORR at Crossover Day 56

Durable ORR at Crossover Day 56 is defined as the proportion of all crossover patients who achieve a complete response (CR) or partial response (PR) at Crossover Day 28 and maintain a CR or PR at Crossover Day 56. Durable ORR at Crossover Day 56 will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] based on the CAS using local investigators' review of aGvHD assessment data. Note that response is relative to the last assessment of aGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a durable responder at Crossover Day 56 if any of the following events occurs:

- Not a responder at Crossover Day 28
- Missing aGvHD assessment at Crossover Day 56
- No CR or PR at Crossover Day 56.
- Additional systemic therapy for aGvHD prior to Crossover Day 56

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Crossover Day 28 to Crossover Day 56 value for the patients who achieved PRs at Crossover Day 28.

2.7 Analysis of secondary efficacy objective(s)

The other secondary efficacy objectives are to:

- Evaluate the two treatment arms with respect to ORR at other time points, e.g. Day 14
- Evaluate the two treatment arms with respect to duration of response (DOR)
- Evaluate the two treatment arms with respect to overall survival (OS)
- Evaluate the two treatment arms with respect to event free survival (EFS)
- Evaluate the two treatment arms with respect to failure free survival (FFS)
- Evaluate the two treatment arms with respect to Non-relapse mortality (NRM)
- Evaluate the two treatment arms with respect to incidence of malignancy relapse/progression (MR)
- Describe cumulative steroid dosing until Day 56 and until EOT in each treatment arm
- Evaluate the two treatment arms with respect to incidence of cGvHD
- Evaluate the two treatment arms with respect to BOR at any time points up to Day 28

All the secondary efficacy endpoint analyses are non-comparative in nature and will be analyzed using the Full Analysis Set (FAS).

2.7.1 Secondary efficacy endpoints

Overall Response Rate at Day 14

Overall Response Rate at Day 14 is defined as the proportion of patients with complete response (CR) or partial response (PR) at Day 14 according to standard criteria [Harris 2016]. This endpoint was reported in the primary analysis, and will not be repeated in this final analysis.

Duration of response (DOR)

Duration of response is defined for patients whose overall response at Day 28 is complete response (CR) or partial response (PR) according to standard criteria [Harris 2016]. The start date is the date of first documented response of CR or PR (i.e., the start date of response), which could be prior to or at Day 28. If it's prior to Day 28, there should not be progression or addition of systemic therapies for aGvHD between the start date of response and Day 28. The end date is defined as the date of progression or the date of addition of systemic therapies for aGvHD on or after Day 28.

Death without prior observation of aGvHD progression and onset of chronic GvHD are considered to be competing risks.

Duration of response will be censored at the last response assessment prior to or at the analysis cut-off date, if no events/competing risk occurred before or at the cut-off date.

Overall survival (OS)

Overall survival is defined as the time from date of randomization to date of death due to any cause. If a patient is not known to have died, then OS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Event Free Survival (EFS)

Event-free survival is defined as the time from the date of randomization to the date of hematologic disease relapse/progression, graft failure or death due to any cause. If a patient is not known to have any event, then EFS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Failure Free Survival (FFS)

Failure-free survival is defined as the time from the date of randomization to date of hematologic disease relapse/progression, non-relapse mortality or addition of new systemic aGvHD treatment.

Non-relapse mortality (NRM)

Non-relapse mortality is defined as the time from date of randomization to date of death not preceded by hematologic disease relapse/progression. Hematologic disease relapse/progression is considered a competing risk for NRM with the date of hematologic disease relapse/progression being the earlier of documented hematologic disease relapse/progression or institution of therapy to treat potential hematologic disease relapse/progression. If a patient is not known to have died or to have relapsed/progressed, then NRM will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Incidence of Malignancy Relapse/Progression (MR)

Malignancy relapse/progression is defined as the time from date of randomization to hematologic malignancy relapse/progression. Deaths not preceded by hematologic malignancy relapse/progression are competing risks. If a patient is not known to have event or competing risks, then MR will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Incidence of cGvHD

cGvHD is defined as the diagnosis of any cGvHD including mild, moderate, severe. Incidence of cGvHD is the time from date of randomization to onset of cGvHD. Deaths without prior onset of cGvHD and hematologic disease relapse/progression are competing risks. If a patient is not known to have event or competing risks, then the incidence of cGvHD will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Best Overall Response

Best overall response (BOR) is defined as proportion of patients who achieved overall response (CR or PR) at any time point up to and including Day 28, and had no additional systemic therapy for aGvHD prior to the time point. This endpoint was reported in the primary analysis, and will not be repeated in this final analysis.

2.7.2 Statistical hypothesis, model, and method of analysis

Overall Response Rate at Day 14

Overall response rate will be calculated based on the local investigators review of aGvHD assessment data. It will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

Duration of response (DOR)

The estimated cumulative incidence rates and 95% confidence intervals at 1, 2, 6, 12, 18 and 24 months will be presented for each treatment group. The cumulative incidence curve will be plotted. DOR will be listed and summarized by treatment group for all patients in the FAS with overall response of CR or PR at Day 28.

Overall survival (OS)

OS will be analyzed according to the randomized treatment group and strata assigned at randomization (aGvHD grade: Grade II vs. III vs. IV). The OS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 1, 2, 6, 12, 18 and 24 month survival estimates and 95% confidence intervals [Brookmeyer and Crowley 1982] will be presented for each treatment group. The hazard ratio for OS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

OS follow-up time will be summarized by treatment group for all patients in FAS. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval.

Event Free Survival (EFS)

EFS will be analyzed according to the randomized treatment group and strata assigned at randomization. The EFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 1, 2, 6, 12, 18 and 24 month survival estimates and 95% confidence intervals [Brookmeyer and Crowley 1982] will be presented for each treatment group. The hazard ratio for EFS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

A sensitivity analysis including aGvHD progression as an event will be performed using the same approach.

Failure Free Survival (FFS)

Cumulative incidence curve for FFS as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for each treatment group, accounting for onset of chronic GvHD as the competing risk.

The FFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curveswill be presented for each treatment group. The hazard ratio for FFS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

In addition, the cumulative incidence of each of the three components considering the other two components as a competing risks will be estimated. Onset of chronic GvHD is considered as a competing risk for all three types of failure. The cumulative incidence curves will be plotted for each treatment group.

A sensitivity analysis including aGvHD progression as an event will be performed using the same approach.

Non-relapse mortality (NRM)

NRM will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization (aGvHD grade: Grade II vs. III vs. IV). The cumulative incidence curve for NRM as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for each treatment group.

As a sensitivity analysis, the cumulative incidence curve for NRM as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for patients with underlying hematologic malignant disease in each treatment group.

Incidence of Malignancy Relapse/Progression (MR)

The cumulative incidence curve for MR and estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for patients with underlying hematologic malignant disease, accounting for NRM as the competing risk.

In addition, the proportion of patients who had hematologic malignancy relapse/progression and its 95% confidence interval at 1, 2, 6, 12, 18 and 24 months will be presented by treatment group for patients with underlying hematologic malignant disease. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

Cumulative steroid dosing until Day 56

The dosing of corticosteroid at baseline was defined as the starting dose at the time of randomization/Day 1. If Day 1 dose was missing, the Day 2 dose would be used.

Overall and weekly cumulative steroid dose for each patient up to Day 56 or discontinuation of randomized treatment will be tabulated. In addition, the relative dose intensity (RDI), by week, will be calculated relative to the starting dose of corticosteroids and categorized as (1) complete reduction where patients are tapered off corticosteroids by D56, (2) $\leq$50% RDI and (3) >50% RDI. The proportion of patients in each category and corresponding 95% confidence intervals will be presented by treatment group. Odds ratio and 95% confidence limits calculated from Fisher's exact test will be also presented for the proportion of patients who are completely tapered off corticosteroids by D56.

In addition, the proportion of patients who were able to reduce any dose or 50% of corticosteroids dose until D56 from baseline will be provided. The proportion of patients in each category and corresponding 95% confidence intervals will be presented by treatment group.

Among those patients who reduced the dose of corticosteroids, the percentage change in dose from baseline to D56, and the maximum of dose reducation during the period will be calculated for each subject. The descriptive statistics will be provided.

Cumulative steroid dosing until EOT

The proportion of patients with complete reduction, where patients are tapered off corticosteroids by EOT, will be provided. [redacted]

[redacted]

Average corticosteroid dose (both original dose and weight-adjusted dose) during the week ending on Days 14, 28, 56, 84, and 168 will also be tabulated and plotted.

Incidence of cGvHD

The cumulative incidence of cGvHD and estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented, accounting for competing risks. The cumulative incidence curve will be plotted for each treatment group.

The listing of cGvHD will be provided. Onset of cGvHD (first event) will summarized. The proportion of subjects with different overall severity of cGvHD will be presented, as well as the proportion of patients with required systemic treatment for cGvHD.

Best Overall Response

Best overall response will be calculated based on the local investigators review of aGvHD assessment data. It will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

2.7.3 Handling of missing values/censoring/discontinuations

Refer to Section 2.5.3

2.8 Safety analyses

All safety analyses will be based on Safety Set, except that the summary of safety data during the crossover treatment phase will be based on Crossover Analysis Set. All listings and tables will be presented by treatment group. For safety evaluations (except for AE) during randomized treatment phase, the last available assessment on or before the date of start of randomized treatment is taken as the "baseline" assessment. For safety evaluations (except for AE) during crossover treatment phase, the last available assessment on or before the date of start of crossover treatment is taken as the "baseline" assessment.

Due to possible crossover from BAT to ruxolitinib arm after Day 28, imbalance in exposure between the two arms is expected. Therefore, safety summaries for the randomized treatment will be performed for the following periods, unless specified:

- Up to Day 31 (the upper bound of the Day 28 visit window);
- Up to the earlier of i) cutoff date, ii) end date of on-randomized-treatment period;

The on-randomized-treatment period is defined in Section 2.1.1.

For the summaries up to cutoff date, some may be presented by adjusting for the duration of randomized treatment period in patient-years where relevant.

2.8.1 Adverse events (AEs)

For reporting of AEs the overall observation period will be divided into mutually exclusive categories, including pre-treatment, on-treatment (randomized or crossover), post-treatment periods as defined in Section 2.1.1.

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

AEs will be summarized by number and percentage of patients having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding. A patient with multiple occurrences of an AE will be counted only once in the respective AE category. A patient with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

In AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on the descending frequency in the ruxolitinib arm.

The following adverse event summaries will be produced by treatment arm; overview of adverse events and deaths, AEs by SOC and PT, summarized by relationship (all AEs and AEs related to study treatment), seriousness (SAEs and non-SAEs), leading to treatment discontinuation, leading to dose interruption/adjustment, and leading to fatal outcome. In addition, a summary of serious adverse events with number of occurrences will be produced (an occurrence is defined as >1 day between start and prior end date of record of same preferred term).

2.8.1.1 AEs adjusted for patient duration of treatment period

In order to account for differences in exposure of the ruxolitinib arm relative to the BAT arm due to crossover from BAT to ruxolitinib after Day 28 visit, incidence rates of adverse events may be presented by adjusting for duration of treatment period in patient-years where relevant.

2.8.1.2 Adverse events of special interest / grouping of AEs

Data analysis of AESIs

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to compound ruxolitinib. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and/or PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad.

For each AESI (selection from the Case Retrieval Sheet), number and percentage of patients with at least one event of the AESI occurring during on-treatment period will be summarized. Summaries of AESIs will be provided by treatment arm, (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose adjustment/interruption, hospitalization, death etc.).

A listing of all grouping levels down to the MedDRA preferred terms used to define each AESI will be generated.

In addition to summarizing infections by CTCAE grade, they will also be summarized using infection severity (protocol Appendix 2) up to Day 31 (the upper bound of the Day 28 visit window), and cutoff date. Summary of infections by SOC, HLGT, HLT and PT up to Day 31, and cutoff date will be generated.

Proportion of patients developing second primary malignancies will be summarized for on-treatment period, post-treatment period and both.

2.8.2 Deaths

Separate summaries for on-treatment and all deaths (including post-treatment death and crossover-period death) will be produced by treatment arm, system organ class and preferred term.

All deaths will be listed and post-treatment deaths will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened patients.

2.8.3 Laboratory data

On analyzing laboratory, data from all sources (central and local laboratories) will be combined. The summaries will include all assessments available for the lab parameter collected no later than end of the on-treatment periods (randomized or crossover).

Grading of laboratory values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher.

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

- Shift tables using CTC grades to compare baseline to the worst on-treatment value
- Shift tables using CTC grades to compare baseline to Day 28

The following listings will be produced for the laboratory data:

- Listings of all laboratory data, with CTC grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.
- Listing of all CTC grade 3 or 4 laboratory toxicities

In addition to the above mentioned tables and listings, other exploratory analyses, for example figures plotting time course of raw or change in laboratory tests over time or box plots may be produced.

The categorical analysis of hepatic lab values will be summarized for both randomizd treatment period and crossover treatment period, displayed in a separate table. Please use the Novartis Hepatotoxicity Guideline 'Update to Hepatic lab values TFLs_released 28Feb2020' for reference. The table of categorical hepatic lab values should now flag 'Combined elevations post-baseline' where 'combined elevations' are 'based on the peak values at any post-baseline time for a subject'.

The following listings will be produced for the laboratory data:

- Listings of hepatic laboratory data, with CTC grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

Not applicable.

2.8.4.2 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: height (cm), weight (kg), body mass index (kg/m2), body temperature (°C), pulse (beats per minute), systolic and diastolic blood pressure (mmHg).

Data handling

Vital signs collected on treatment will be summarized. Values measured outside of on treatment period will be flagged in the listings.

Data analysis

For analysis of vital signs the clinically notable vital sign criteria are provided in Table 2-5 below.

Table 2-5 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Weight (kg)	increase > 10% from Baseline	decrease > 10% from Baseline
Systolic blood pressure (mmHg)	>=180 with increase from baseline of >=20	<=90 with decrease from baseline of >=20
Diastolic blood pressure (mmHg)	>=105 with increase from baseline of >=15	<=50 with decrease from baseline of >=15

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Pulse rate (bpm)	>=100 with increase from baseline of >25%	<=50 with decrease from baseline of > 25%
Body temperature	>= 39.1	-

The number and percentage of patients with notable vital sign values (high/low) will be presented by treatment arm.

A listing of all vital sign assessments will be produced by treatment arm and notable values will be flagged. In the listing, the assessments collected outside of on-treatment period will be flagged.

Additional Analyses

Time to first occurrence of grade 3 infection

Time to first occurrence of infection is defined as time from start of study treatment to the date of first occurrence of grade 3 infection severity per protocol Appendix 2, i.e. time in days is calculated as (start date of first occurrence of infection) – (start of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be **the earliest** of the following dates:

- death date
- end date of on-treatment period (randomized or crossover)
- data cut-off date
- withdrawal of informed consent date

Failure curves (ascending Kaplan-Meier curves) will be constructed by treatment arm. Median together with 95% confidence interval as well as 25th percentile and 75th percentile will be presented for each treatment arm.

A sensitivity analysis will be conducted considering deaths or end of treatment phase (randomized or crossover) without prior infection as competing risks. Cumulative incidence curve for time to grade 3 infection as well as estimates at 1, 2 and 6 months with 95% confidence intervals will be presented for each treatment group.

In addition, the median time to occurrence for the subset of patients who experienced infection will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.

2.9 Pharmacokinetic endpoints

The pharmacokinetic endpoints were reported in the primary analysis, and has not been reported in the second analysis. Further results will be presented in the full analysis.

Pharmacokinetic Analysis Set (PAS) will be used in all pharmacokinetic data analysis and PK summary statistics.

Plasma samples for extensive PK will be taken at Day 1 (start of treatment), at Day 7 (week 1) for the first 25 adult patients and all adolescents to characterize the PK after first dose, and at steady state by non-compartmental analysis. Additional PK samples will be taken at later visits for all patients to characterize exposure-efficacy, exposure-safety [redacted] as data allows. Concentrations will be expressed in mass per volume units.

PK parameters

The PK parameters that will be determined are shown in Table 2-6. The PK parameters of ruxolitinib will be calculated from the extensive PK data based on the non-compartmental methods using Phoenix WinNonlin® (Pharsight, Mountain View, CA) software. Additional PK parameters may be estimated as needed.

Table 2-6 Non-compartmental PK parameters for ruxolitinib

AUClast	The AUC from time zero to the last measurable concentration sampling time (tlast) (ng x hr/mL)
AUCinf	The AUC from time zero to infinity (ng x hr/mL)
AUCtau	The AUC calculated to the end of a dosing interval (12 hr) at steady-state (ng x hr/mL)
Cmax	The maximum (peak) observed plasma, blood, serum, or other body fluid drug concentration after single dose administration (ng/mL)
Ctrough	The observed plasma concentration obtained prior to administration of the next dose (pre-dose concentration) (mass x volume-1)
Tmax	The time to reach maximum (peak) plasma, blood, serum, or other body fluid drug concentration after single dose administration (thr)
Lambda_z	Smallest (slowest) disposition (hybrid) rate constant (hr^{-1}) may also be used for terminal elimination rate constant (hr^{-1})
T1/2	The elimination half-life associated with the terminal slope (λz) of a semi logarithmic concentration-time curve (hr).
CL/F	The total body clearance of drug from the plasma (L/hr)
Vz/F	The apparent volume of distribution during terminal phase (associated with λz) (L)
Racc	Accumulation ratio (AUC at steady state/AUC Day 1)

Descriptive statistics (n, arithmetic mean, CV% mean, standard deviation (SD), median, geometric mean, CV% geo-mean, minimum and maximum) will be presented for the ruxolitinib arm for Pharmacokinetic Analysis Set for all PK parameters defined in Table 2-6 except Tmax, where only n, median, minimum and maximum will be presented.

All individual PK parameters will be listed for the ruxolitinib arm using the Full Analysis Set.

The potential impact of severity of GI GvHD on ruxolitinib pharmacokinetic parameters will be explored by summarizing PK parameters by baseline lower GI stage and producing box plots of AUCinf, AUCtau and Cmax by lower GI stage.

PK parameters will also be summarized by patients who took concomitant CYP3A4 inhibitors versus those who didn't. Patients are defined as having taken concomitant CYP3A4 inhibitors if based on the concomitant medication summary they have taken CYP3A4 inhibitors on the day of or on the day before PK samples are taken. In addition to the summary tables box plots of Cmax, AUCinf and AUCtau will be produced by concomitant CYP3A4 inhibitor use. This will be repeated for Day 1 and Day 7 PK parameters. The list of CYP3A4 inhibitors will be documented in Data Handling Plan.

[redacted]

PK concentrations

Descriptive statistics (n, m (number of non-zero concentrations), arithmetic mean, CV% mean, SD, median, geometric mean, CV% geo-mean, minimum and maximum) for ruxolitinib concentration by lower GI stage and overall will be presented at each scheduled time point for the ruxolitinib arm for the Pharmacokinetic Analysis Set.

Individual concentration-time profiles for ruxolitinib concentrations with median will be displayed graphically for the ruxolitinib arm for Full Analysis Set on the semi-log view. In addition, the mean (+/- SD) and geometric mean concentration-time profiles for ruxolitinib by treatment over time will be displayed graphically for Pharmacokinetic Analysis Set on the linear and semi-log view.

All individual plasma ruxolitinib concentration data will be listed for the ruxolitinib arm for the Full Analysis Set.

Crossover concentrations will be included in all analyses on concentrations.

Handling of PK data below LLOQ or missing

All concentration values below the lower limit of quantitation (LLOQ) are set to zero by the Bioanalyst, and will be displayed in the listings as zero and flagged. LLOQ values will be treated as zero in any calculations of summary statistics, and treated as missing for the calculation of the geometric means and their CV%. The number of non-zero concentrations will also be reported in the summary statistics.

Missing values for any PK data will not be imputed and will be treated as missing.

2.10 PD and PK/PD analyses

Analysis of relationship between ruxolitinib exposure and efficacy/safety endpoints

This analysis will be described and reported separately to the CSR. The following are objectives for exposure-response analysis:

- Characterize the exposure-efficacy relationship of ruxolitinib in terms of exposure and efficacy response, with efficacy response defined as (overall response rate at Day 28 and durable response at Day 56, duration of response, overall survival at 6 months and any other relevant endpoints, and exposure defined plasmaconcentration, PK parameter or dose, as appropriate)
- Characterize the exposure-safety relationship of ruxolitinib in terms of exposure and safety response, where safety response is defined as various categorizations of AEs (frequency of;severity of AEs, AEs of interest to be defined prior to analysis or laboratory parameters).
- Average steady-state exposures and/or other PK parameters for the population will becomputed by the POP PK model accounting for dose modifications or dose interruptions up to the day prior to the day of assessments. Population PK derived parameters will be used for exposure-response analysis by appropriate methods

For duration of response and overall survival, if the primary endpoint is significant and sufficient events have accrued (i.e., at analysis time points after the primary analysis), a Cox regression model with appropriate patient demographic and prognostic factors as covariates and the log-trough level as a time dependent covariate will be fitted if appropriate. To account for dose adjustments between trough sampling time points the trough concentration will be adjusted by dividing by the actual dose prior to the trough in question and multiplying by the arithmetic mean dose since the last trough sample or start of dosing as applicable. The calculation of the arithmetic mean dose should count any dose interruptions as a zero dose for the days that no dose was given. Goodness of fit of the model will also be examined. The survival time should be calculated from the time of first dose rather than the time of randomization. The hazard ratio and 95% confidence interval for a two-fold increase in exposure will be displayed if appropriate. Kaplan-Meier curves may be used to summarize the data based on relevant quantiles of PK.

For incidence of specific AEs, Day 28 response (if the primary endpoint is significant) and Day 56 durable response (if the primary and key secondary endpoints are significant), logistic regression models may be used including log-average trough concentration and other demographic and prognostic covariates in the model as appropriate. The trough concentration will be similarly adjusted to account for dose adjustments as described above. The average trough concentration included in the model for Day 28 response will be the arithmetic mean of the Day 7, Day 14, Day 21 and Day 28 adjusted trough concentration. A similar average trough concentration will be calculated for the Day 56 response except the Day 42 and Day 56 adjusted concentration will be double weighted compared to the others because they are representative of concentration over 2 weeks rather than 1. For analysis of adverse events the last trough

concentration prior to the AE will be included in the model. Other exposure measures may be considered if appropriate. The odds ratio and its 95% confidence interval for a two-fold increase in exposure will be displayed as appropriate.

2.11 Patient-reported outcomes

The FACT-BMT along with the EQ-5D-5L will be used to collect data on the patient's disease-related symptoms and health-related quality of life. Responses to the FACT-BMT and EQ-5D-5L will be generated in accordance with the respective scoring manual. Further details on the scoring of FACT-BMT and EQ-5D are given in Appendix 5.4.3.1 and Appendix 5.4.3.2.

These PRO instruments are planned to be administered on randomization day and every week during the first 2 months, and every 4 weeks thereafter until the end of treatment.

The baseline is defined as the last PRO assessment prior to or on the treatment start date .

FACT-BMT will not be used in patients under age 18 years.

Descriptive statistics (mean, standard deviation, median, Q1, Q3, minimum, and maximum) will be used to summarize the scored scales at each scheduled assessment time point for the FACT-BMT and EQ-5D using FAS. Additionally, change from baseline in the scores at the time of each assessment will be summarized. Patients with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

Missing items data in a scale will be handled based on each instrument manual. No imputation will be applied if the total or subscale scores are missing at a visit.

Subscale scores from the FACT-BMT and EQ-5D-5L will be displayed as mean profiles for each study arm, presented over time using time windows as described in Section 2.1.

[redacted]

[redacted]

2.14 Interim analysis

No formal interim analysis is planned for this trial.

Early PK analysis

To compare the exposure of 10 mg BID in the SR aGvHD population to the known exposure in MPN patients, the early extensive ruxolitinib PK data on the first 25 adult patients (including any adolescent patients randomized at that time) will be explored once available, also in the context of concomitant medications. However, there will be no comparison between the two treatment arms because PK data are not collected in BAT patients.

DMC safety data analysis

DMC will be instituted in this study and will review the safety data as outlined according the DMC charter. A separated set of SAP and TFL shells will be provided to detail the analysis.

3 Sample size calculation

The study with a total of 308 patients and 1:1 randomization (ruxolitinib vs. BAT) stratified on aGvHD grade (Grade II vs. Grade III vs. Grade IV) has 90% power to test for the primary endpoint (ORR at Day 28) and approximately 90% power to test for the key secondary endpoint (durable ORR at Day 56). The family wise α-level will be controlled at 0.025 overall for the two comparisons. Specifically, this study will claim to have achieved the efficacy objective when the primary endpoint ORR at Day 28 shows a significant treatment effect at one-sided $\alpha = 0.025$. Conditional on significance of the primary endpoint, the key secondary endpoint durable ORR at Day 56 will be tested at one-sided $\alpha = 0.025$.

3.1 Primary analysis

The sample size calculation is based on the primary variable ORR at Day 28. The hypotheses to be tested and details of the testing strategy are described in Section 2.5.2 and Section 3.

Based on [Martin P. et al BBMT 2012], the ORR at Day 28 in the BAT arm is expected to be 58%. The stratum specific rates (Grade II 69%, Grade III 59%, Grade IV 50%) are obtained

assuming that the ratio of aGvHD Grade II: III: IV is 0.2:0.4:0.4. It is expected that treatment with ruxolitinib will result in an 18% increase in the ORR, i.e., an expected odds ratio of 2.25 (which corresponds to an increase in ORR to 75%). Power for the CMH test, stratifying on aGvHD grade, was calculated using software package East V6. In order to ensure 90% power a total sample size of 308 patients are needed.

With a sample size of 154 patients in each treatment arm, an observed odds ratio greater than or equal to 1.63 would achieve statistical significance for the primary endpoint. Assuming that the observed response rates in Grades II/III/IV in BAT arm are 69%/59%/50% (overall 57%), observed response rates ≥ 78%/70%/62% (overall 68%) in the ruxolitinib arm would achieve statistical significance.

3.2 Power for analysis of key secondary variables

Durable ORR at Day 56, as the key secondary variable, will be formally statistically tested, provided that the primary endpoint ORR at Day 28 is statistically significant. The hypotheses to be tested and details of the testing strategy are provided in Section 2.6.2. Based on [Van Groningen 2016], the durable ORR at Day 56 in the BAT arm is expected to be approximately 35%. The stratum specific rates (Grade II 45%, Grade III 36%, Grade IV 30%) are obtained assuming that the ratio of aGvHD Grade II: III: IV is 0.2:0.4:0.4. It is expected that treatment with ruxolitinib will result in a 20% increase in the ORR, i.e., an expected odds ratio of 2.25 (which corresponds to an increase in durable ORR to 55%). With these assumptions and sample size of 308 patients, the power for the key secondary endpoint is at least 90%.

With sample size 154 patients in each treatment arm, an observed odds ratio greater than or equal to 1.59 for durable ORR at Day 56 would achieve statistical significance. Assuming that the observed durable response rates in Grades II/III/IV in BAT arm are 45%/36%/30% (overall 35%), observed durable response rates ≥ 57%/47%/41% (overall 47%) in the ruxolitinib arm would achieve statistical significance.

4 Change to protocol specified analyses

Compared to the protocol version 00, ‘discontinuation from randomized treatment’ has been dropped as reason. Since some BATs have fixed duration which is less than 8 weeks, to avoid bias, discontinuation from randomized treatment will not be used to disqualify the response or durable response. Also duration of response will not be censored based on treatment discontinuation.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

The following rule should be used for the imputation of the dose end date for a given study treatment component:

Scenario 1: If the dose end date is completely missing and there is no EOT page and no death date, the patient is considered as on-going:

The patient should be treated as on-going and the cut-off date should be used as the dose end date.

Scenario 1 should not be applied for final CSR. All patients should have EOT page complete before the Database lock for Final CSR

Scenario 2: If the dose end date is completely or partially missing and the EOT page is available:

Please note that date of assessment on EOT eCRF might be very different from last date of dose.

Case 1: The dose end date is completely missing, and the EOT completion date is complete, then this latter date should be used.

Case 2: Only Year(yyyy) of the dose end date is available and yyyy < the year of EOT date:

Use Dec31yyyy

Case 3: Only Year(yyyy) of the dose end date is available and yyyy = the year of EOT date:

Use EOT date

Case 4: Both Year(yyyy) and Month (mm) are available for dose end date, and yyyy = the year of EOT date and mm < the month of EOT date:

Use last day of the Month (mm)

All other cases should be considered as a data issue and the statistician should contact the data manager of the study.

After imputation, compare the imputed date with start date of treatment, if the imputed date is < start date of treatment:

Use the treatment start date

Patients with missing start dates are to be considered missing for all study treatment component related calculations and no imputation will be made. If start date is missing then end-date should not be imputed.

5.1.2 AE, ConMeds and safety assessment date imputation

Table 5-1 Imputation of start dates (AE, CM) and assessments (LB, EG, VS)

Missing Element	Rule
day, month, and year	• No imputation will be done for completely missing dates

day, month	• If available year = year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY
	  ○ Else set start date = study treatment start date.
	• If available year > year of study treatment start date then 01JanYYYY

Missing Element	Rule
	• If available year < year of study treatment start date then 01JulYYYY
day	• If available month and year = month and year of study treatment start date then o If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY. o Else set start date = study treatment start date. • If available month and year > month and year of study treatment start date then 01MONYYYY • If available month and year < month year of study treatment start date then 15MONYYYY

Table 5-2 Imputation of end dates (AE, CM)

Missing Element	Rule (* if end date of the on-treatment period not > (death date, cut-off date, withdrawal of consent date))
day, month, and year	• Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	• If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period *
day	• If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

Any AEs and ConMeds with partial/missing dates will be displayed as such in the data listings.

Any AEs and ConMeds which are continuing as per data cut-off will be shown as 'ongoing' rather than the end date provided.

The above imputations are only used for analyses of time to and duration of AEs and concomitant medications.

5.1.2.1 Other imputations

Incomplete date of initial diagnosis of aGvHD

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

5.2 AEs coding/grading

Adverse events are coded using the Medical dictionary for regulatory activities (MedDRA) terminology.

The latest available MedDRA version at the time of the analyses should be used. The MedDRA version used should be specified in the footnote of relevant tables.

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 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).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTCAE grades are given in Novartis internal criteria for CTCAE grading of laboratory parameters. The latest available version of the document based on the underlying CTCAE version 4.03 at the time of analysis will be used. For laboratory tests where grades are not defined by CTCAE v4.03, results will be graded by the low/normal/high (or other project-specific ranges, if more suitable) classifications based on laboratory normal ranges.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. CTCAE Grade 5 is not defined for laboratory values. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

Imputation Rules

CTC grading for blood differentials is based on absolute values. However, this data may not be reported as absolute counts but rather as percentage of WBC.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

The following rules will be applied to derive the WBC differential counts when only percentages are available for a xxx differential

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

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4]

In order to apply the above formula, albumin values in g/L will be converted to g/dL by multiplying by 0.1), calcium values in mmol/L will be converted to mg/dL by dividing by 0.2495. For calculation of laboratory CTC grades 0 and 1, the normal range for derived corrected calcium is set to the same limits (in mg/dL) as for calcium.

CTC grades for the derived absolute WBC differential counts (neutrophils, lymphocytes) and corrected calcium will be assigned as described above for grading.

5.4 Statistical models

5.4.1 Primary analysis

The null hypothesis of equality of response rate in the two treatment arm will be tested against one-sided alternative. The statistical hypotheses are:

$$H_0: ORR_{rux} \leq ORR_{BAT} \quad \text{vs.} \quad H_1: ORR_{rux} > ORR_{BAT}$$

where ORR_{rux} is the probability of response in ruxolitinib and ORR_{BAT} is the probability of response in BAT.

The Cochran-Mantel-Haenszel chi-square test X^2_{CMH} (implemented again via SAS procedure FREQ with CMH option in the TABLES statement) will be used to test the difference in response rates between the treatment arms. The p-value corresponding to the CMH test for "general association" will be used which follows a Chi-square distribution with one degree of freedom.

If the sampling assumptions for chi-square test is not met, the exact Cochran-Mantel-Haenszel test will be used (implemented via SAS procedure MULTTEST). The test is performed by running a stratified version of the Cochran-Armitage permutation test [Armitage et al. 1969]. In studies with stratified randomization, the chi-square approximation is considered appropriate for the X^2_{CMH} statistics if the rule of Mantel and Fleiss [Mantel and Fleiss 1980] is satisfied.

Logistic Regression

Odds ratio will be used as a measure of association between treatment and response. The odds ratio will be derived from the logistic regression model (implemented using SAS procedure LOGISTIC, with treatment specified as an explanatory variable in the CLASS statement) which allows for including not only the stratification factor but also for adjustments for other covariates (both categorical and continuous). The odds ratio will be presented with 95% Wald confidence limits.

In cases where an exact test has been used to compare response rates, the odds ratio should be determined using exact logistic regression, and the odds ratio presented with exact 95% confidence limits. In these cases, SAS PROC LOGISTIC with EXACTONLY option will be used.

Confidence interval for response rate

Responses will be summarized in terms of percentage rates with $100(1-\alpha)\%$ confidence interval using exact binomial confidence interval (implemented using SAS procedure FREQ with EXACT statement for one-way table [Clopper and Pearson 1934].

Definition of new or additional systemic therapy for aGvHD

The data source to search for the new or additional systemic therapy would be from eCRFs of Prior and Concomitant Medication (CONMED) (with subcategory as treatment for aGvHD) and Dosage Administration Record (DAR).

Any of the following therapies represent new or additional systemic therapy for aGvHD:

1. Any new CNI therapy being initiated as 'treatment for aGvHD' after the baseline as recorded either on DAR or on CONMED, and never received prior to or at baseline
2. Any other systemic therapy (excluding CNI or systemic corticosteroid (methylprednisolone, prednisone and prednisolone)) being started as 'treatment for aGvHD' after the baseline and recorded on CONMED. Note: the therapy may have been taken prior to baseline but not at baseline.
3. Additional BAT at or after start of the initial BAT as recorded on DAR
4. Treatment with ruxolitinib after cross-over from BAT as recorded on DAR

5.4.2 Key secondary analysis

Same instructions as in Section 5.4.1.

5.4.3 Other secondary analysis

5.4.3.1 FACT-BMT

For FACT-BMT the subscale scores, the FACT-BMT total score and the FACT-BMT Trial Outcome Index (TOI) will be calculated.

The subscales are physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), functional well-being (FWB) and bone marrow transplant subscale (BMTS). For each subscale the corresponding score will be calculated based on the item response of the answered question according to the FACT-BMT Scoring Guide (Version 4). The scoring guide identifies those items that must be reversed before being added to obtain subscale totals. Negatively stated items are reversed by subtracting the response from "4". After reversing proper items, all subscale items are summed to a total, which is the subscale score.

If there are missing items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done by using the formula below:

Prorated subscale score

= [Sum of item scores] x [N of items in subscale]/[N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the relevant items were answered. Note that not all items (questions) are used to calculate the scores (see Scoring Guide, version 4). For all the scores the higher the score is the better the QOL is.

The FACT-BMT Trial Outcome Index (TOI, score range: 0-96) is calculated as

$$\underset{(\textbf{PWB score})}{\_\_\_\_\_} + \underset{(\textbf{FWB score})}{\_\_\_\_\_} + \underset{(\textbf{BMTS score})}{\_\_\_\_\_} = \_\_\_\_\_ = \textbf{FACT-BMT TOI}$$

The FACT-BMT total score (score range: 0-148) is then calculated as the sum of all un-weighted subscale scores:

(PWB score) + (SWB score) + (EWB score) + (FWB score) + (BMTS score) = **FACT-BMT Total score**

5.4.3.2 EQ-5D-5L

EQ-5D-5L consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Also the EQ-VAS is included. Each dimension has 5 levels (1-5).

The data were reported for all 5 dimensions. The proportion of reported problems for each level were reported. The changes from baseline were defined with a shift table. The below screenshot is from the Excel conversion document obtained from the EuroQol webpage on 5L Value Sets (http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html). This UK value will be used to convert the "health state" to an "index value" for all patients in all countries. For example, if the health state is 11125, then the index value would be .316 and so on.

EQ-5D-5L_Crosswalk_Value_Sets.xls [Read-Only] [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J	K	L
1		Health state	Denmark	France	Germany	Japan	Netherlands	Spain	Thailand	UK	US	Zimbabwe
2		5L profile	Denma	Franc	Germa	Jap	Netherlan	Spa	Thaila	U	U	Zimbab
3		11111	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.900
4		11112	0.856	0.929	0.999	0.829	0.845	0.932	0.814	0.879	0.876	0.864
5		11113	0.818	0.910	0.999	0.785	0.805	0.914	0.766	0.848	0.844	0.854
6		11114	0.671	0.769	0.809	0.761	0.592	0.731	0.660	0.635	0.700	0.792
7		11115	0.519	0.622	0.611	0.736	0.370	0.541	0.549	0.414	0.550	0.727
8		11121	0.859	0.910	0.910	0.814	0.874	0.910	0.780	0.837	0.861	0.846
9		11122	0.787	0.839	0.909	0.740	0.765	0.857	0.723	0.768	0.820	0.810
10		11123	0.768	0.820	0.909	0.721	0.736	0.843	0.708	0.750	0.809	0.800
11		11124	0.622	0.679	0.719	0.697	0.523	0.660	0.602	0.537	0.669	0.738
12		11125	0.469	0.532	0.521	0.672	0.301	0.470	0.491	0.316	0.524	0.673
13		11131	0.824	0.888	0.887	0.768	0.843	0.887	0.726	0.796	0.827	0.833
14		11132	0.770	0.817	0.887	0.718	0.745	0.838	0.701	0.740	0.806	0.797
15		11133	0.756	0.798	0.887	0.705	0.719	0.825	0.694	0.725	0.800	0.787
16		11134	0.609	0.657	0.697	0.681	0.506	0.642	0.588	0.512	0.661	0.725
17		11135	0.457	0.510	0.499	0.656	0.284	0.452	0.477	0.291	0.517	0.660

6 Reference

1. Anderson JR, Cain KC and Gelber RD (1983). Analysis of survival by tumor response. Journal of Clinical Oncology, 1(11), 710-719.
2. Armitage P, McPherson CK and Rowe BC (1969). Repeated significance tests on accumulating data. J .R. Statist. Soc. A, 132, 232-244.
3. Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41.
4. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrical, 26, 404-413.
5. FDA briefing Document Oncologic Drugs Advisory Committee Meeting, July 24, 2012, Evaluation of radiologic review of progression-free survival in Non-hematologic malignancies. http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/oncologicdrugsadvisorycommittee/ucm311141.pdf
6. Harris AC, Young R, Devine S, Hogan WJ, Ayuk F, Bunworasate U, Chanswangphuwana C, Efebera YA, Holler E, Litzow M, Ordemann R, Qayed M, Renteria AS, Reshef R, Wölfl M, Chen YB, Goldstein S, Jagasia M, Locatelli F, Mielke S, Porter D, Schechter T, Shekhovtsova Z, Ferrara JL, Levine JE. (2016) International, Multicenter Standardization of Acute Graft-versus-Host Disease Clinical Data Collection: A Report from the Mount Sinai Acute GVHD International Consortium. Biol Blood Marrow Transplant; 22(1): 4-10.
7. Korhonen P, Laird NM and Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC study (1999). Statistics in Medicine, 18, 2879-2897.
8. Mantel, N. and Joseph L Fleisś (1980). Minimum expected cell size requirements for the Mantel-Haenszel one degree-of-freedom Chi-Square Test and a related rapid procedure. American Journal of Epidemiology, 112, 129-134.
9. Martin PJ, Rizzo JD, Wingard JR, et al (2012) (2012) First- and second-line systemic treatment of acute graft-versus-host disease: recommendations of the American Society of Blood and Marrow Transplantation. Biol Blood Marrow Transplant; 18(8): 1150-63.
10. Robins JM and Tsiatis AA (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models (RPSFTM). Commun. Statist.-Theory Meth., 20(8), 2609-31.
11. Van Groningen, LF, Liefferink AM, de Haan AF, Schaap NP, Donnelly JP, Blijlevens NM, van der Velden WJ. (2016) Combination Therapy with Inolimomab and Etanercept for Severe Steroid-Refractory Acute Graft-versus-Host Disease Biol Blood Marrow Transplant. 22(1):179-82.

NOVARTIS

Clinical Development

INC424/ruxolitinib/JAKAVI

CINC424C2301 / NCT02913261

A phase III randomized open-label multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease after allogeneic stem cell transplantation

Statistical Analysis Plan (SAP)

Author: Trial Statistician PPD ;

Document type: SAP Documentation

Document status: Final

Release date: 4-Mar-2020

Number of pages: 56

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)

Table of contents

List of abbreviations

AE	Adverse event
AESI	Adverse Event of Special Interest
aGvHD	Acute Graft vs. Host Disease
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
BAT	Best Available Therapy
bid	bis in diem/twice a day
BMI	Body Mass Index
BOR	Best Overall Response
BSA	Body Surface Area
cGvHD	chronic Graft vs. Host Disease
CIBMTR	Center for International Blood and Marrow Transplant Research
CMH	Cochran–Mantel–Haenszel
CMV	Cytomegalovirus
CR	Complete Response
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DAR	Dose Administration Record
DMC	Data Monitoring Committee
eCRF	Electronic Case Report Form
EFS	Event-Free Survival
FACT-BMT	Functional Assessment of Cancer Therapy - Bone Marrow Transplantation
FAS	Full Analysis Set
FFS	Failure-Free Survival
GvHD	Graft vs. Host Disease
HCT	Hematopoietic Cell Transplantation
IRT	Interactive Response Technology that includes Interactive Voice Response System and Interactive Web Response System
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NRM	Non Relapse Mortality
ORR	Overall Response Rate
OS	Overall Survival
PAS	Pharmacokinetic analysis set
PD	Pharmacodynamics
PK	Pharmacokinetics
PPS	Per-Protocol Set
PR	Partial response
PRO	Patient-reported Outcomes
QoL	Quality of Life
RAP	Report and Analysis Process

RPSFT	Rank-Preserving Structure Failure Time
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
WBC	White Blood Cells
SOC	System Organ Class
TFLs	Tables, Figures, Listings
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes all planned analyses for the Clinical Study Report (CSR) of study CINC424C2301, a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease (aGvHD Grade II-IV) after allogeneic stem cell transplantation.

The content of this SAP is based on protocol CINC424C2301 version 02. All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock of the study data.

1.1 Study design

This is a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease after allogeneic stem cell transplantation. Approximately 308 patients will be randomized to one of the following treatment arms in 1:1 ratio:

- Ruxolitinib
- BAT

Randomization will be stratified by the aGvHD grade (Grade II vs. Grade III vs. Grade IV).

Overall response rate (ORR) at Day 28, as assessed by local investigators' review of aGvHD response and using standard criteria [Harris 2016], is the primary endpoint in this study. Durable overall response rate at Day 56 is the key secondary endpoint.

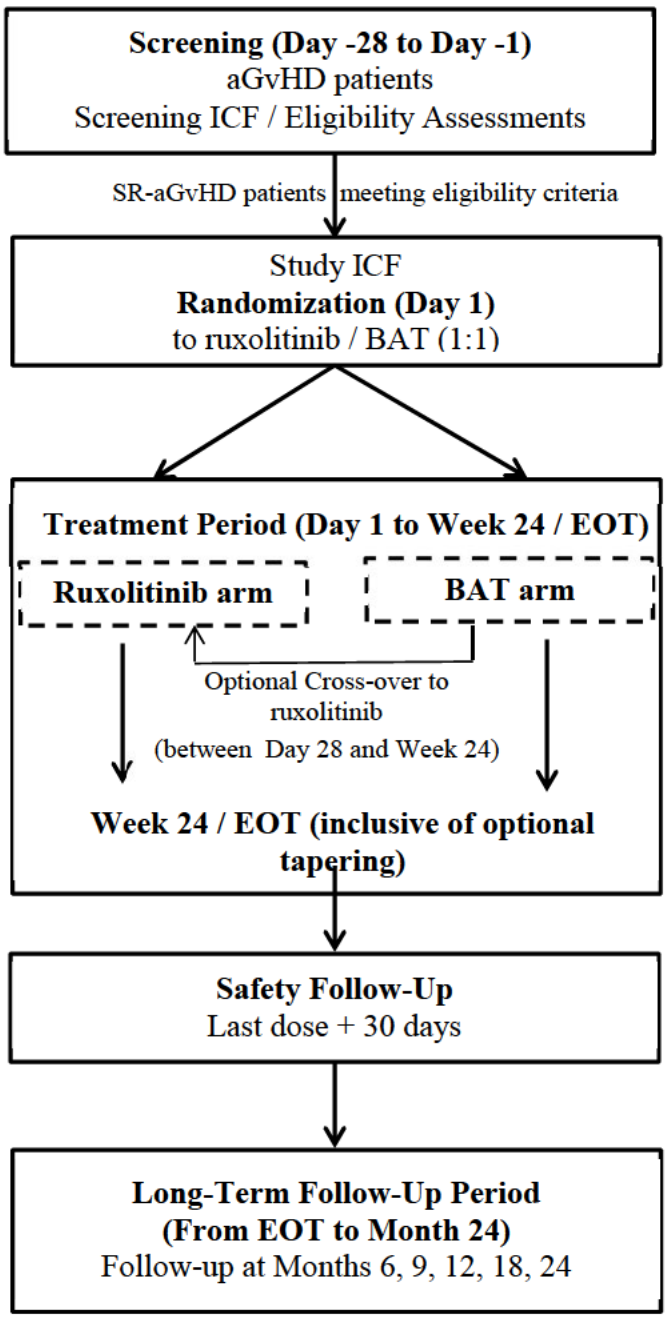
The primary analysis including the analysis on primary and key secondary endpoints was performed with the cut-off date as 25-Jul-2019 after all patients have completed their Day 56 visit or have discontinued study. The primary analysis data were summarized in the primary clinical study report (CSR).

Further analyses on secondary endpoints will be performed when all patients have completed 6 months treatment or discontinued from study participation earlier.

The final analysis will occur once all patients have completed the study (up to 24 months from randomization). All available data from all patients up to end of study, inclusive of OS, will be reported in a final CSR.

No formal interim efficacy analysis is planned in this study.

Figure 1-1 Schematic Study Design



1.2 Study objectives and endpoints

Table 1-1 Objectives and related endpoints

Objective	Endpoint	Analysis
Primary		**Refer to Section 2.5**
To compare the efficacy of ruxolitinib vs. Investigator's choice Best Available Therapy (BAT) in patients with Grade II-IV SR-aGvHD assessed by Overall Response Rate (ORR) at Day 28	Overall response rate (ORR) at Day 28 after randomization, defined as the proportion of patients in each arm demonstrating a complete response (CR) or partial response (PR) without requirement for additional systemic therapies for an earlier progression, mixed response or non-response. Scoring of response will be relative to the organ stage at the time of randomization.	CMH analyses, stratified by aGvHD grade
Key secondary endpoint		**Refer to Section 2.6**
To compare the rate of durable ORR at Day 56 between ruxolitinib and BAT	Proportion of all patients in each arm who achieve a complete response (CR) or partial response (PR) at Day 28 and maintain a CR or PR at Day 56.	CMH analyses, stratified by aGvHD grade
Other secondary endpoints		**Refer to Sections 2.7, 2.8, 2.9, 2.10, 2.11**
To estimate ORR at Day 14	Proportion of patients who achieved ORR (CR+PR) at Day 14.	
To assess Duration of response	Duration of response (DOR) is assessed for responders only and is defined as the time from first response until aGvHD progression or the date of additional systemic therapies for aGvHD. Onset of chronic GvHD, or death without prior observation of aGvHD progression are considered as competing risks.	

Objective	Endpoint	Analysis
To assess the cumulative steroid dose until Day 56	Weekly cumulative steroid dose for each patient up to Day 56 or end of treatment will be calculated.	
To assess Overall Survival (OS)	Overall survival, defined as the time from the date of randomization to the date of death due to any cause.	Probability of overall survival at Months 1, 2, 6, 12, 18 & 24 will be estimated from the Kaplan-Meier curves for each arm.
To assess Event-Free Survival (EFS)	Event-free survival, defined as the time from the date of randomization to the date of hematologic disease relapse/progression, graft failure, or death due to any cause.	Probability of EFS at Months 1, 2, 6, 12, 18 & 24 will be estimated from the Kaplan-Meier curves for each arm.
To assess Failure-Free Survival (FFS)	Failure-free survival, defined as the time from the date of randomization to date of hematologic disease relapse/progression, non-relapse mortality, or addition of new systemic aGvHD treatment.	Cumulative incidence of FFS at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering each event as a competing risk for the other two. Onset of chronic GvHD is considered as a competing risk.
To assess Non Relapse Mortality (NRM)	Non-relapse mortality (NRM), defined as the time from date of randomization to date of death not preceded by hematologic disease relapse/progression.	Cumulative incidence of NRM at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering hematologic disease relapse/progression as competing events.
To assess incidence of Malignancy Relapse/Progression (MR)	Malignancy Relapse/Progression (MR) (refer to protocol Appendix 4), defined as the time from date of randomization to hematologic malignancy relapse/progression. Calculated for patients with underlying hematologic malignant disease.	Cumulative incidence of MR at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering deaths not preceded by hematologic malignancy relapse/progression as competing events.
To measure the incidence of cGvHD	cGvHD, defined as the diagnosis of any cGvHD including mild, moderate, severe.	Cumulative incidence of cGvHD at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering hematologic disease relapse/progression and death without prior cGvHD as competing events.
To estimate the rate of Best Overall Response (BOR)	Proportion of patients who achieve OR (CR+PR) at any time point up to and including Day 28 and	

Objective	Endpoint	Analysis
	before the start of additional systemic therapy for aGvHD.	
To assess Pharmacokinetics (PK) of ruxolitinib in SR-aGvHD patients	Pharmacokinetic parameters of ruxolitinib after a single dose and at steady state. Cmax, AUClast, AUCinf, Ctrough, Racc and AUCtau. Other PK parameters are CL/F, Vz/F, Tmax and T1/2.	
To assess exposure-response relationship of ruxolitinib in SR-aGvHD	Pharmacokinetics (exposure) and efficacy (ORR, OS, or other relevant endpoints) relationship. Pharmacokinetics (exposure) and safety (AEs) relationship.	
To evaluate changes in Patient Reported Outcomes (PROs)	Change in FACT-BMT from baseline to each visit where measured. Change in EQ-5D-5L from baseline to each visit where measured.	
To evaluate the safety of ruxolitinib and Best Available Therapy	Safety and tolerability including myelosuppression, infections, and bleeding will be assessed by monitoring the frequency, duration and severity of Adverse Events including occurrence of any second primary malignancies, infections, by performing physical exams, and evaluating changes in vital signs from baseline, routine serum chemistry, hematology results and coagulation profile.	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	

Objective	Endpoint	Analysis
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
	[redacted]	
	[redacted]	

2 Statistical methods

2.1 Data analysis general information

The final analysis will be performed by Novartis. SAS version 9.4 or later and/or R version 3.0.2 or later will be used to perform all data analyses and to generate tables, figures and listings.

Data included in the analysis

The analysis cut-off date for the second analysis of study data will be established after all randomized patients have completed approximately 6 months treatment or have discontinued treatment period or have discontinued study. All statistical analyses will be performed using all data collected in the database up to the data cutoff date. All data with an assessment date or event start date (e.g. vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'ongoing'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these events, the end date will not be imputed and therefore will not appear in the listings.

The analysis cutoff date for the final analysis of study data will be established when all patients have completed the study.

General analysis conventions

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

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

Quantitative data (e.g., age, body weight, etc.) will be summarized by appropriate

descriptive statistics (i.e. mean, standard deviation, median, minimum, and maximum) by treatment group.

2.1.1 General definitions

Investigational drug and study treatment

Investigational drug, will refer to the ruxolitinib only. Whereas, ***study treatment*** will refer to ruxolitinib and BAT.

Randomized treatment will refer to the study treatment received during the randomized treatment period. Up to Day 28 visit, more than one BAT regimen may be initiated as study

treatment. These regimens reported on the Dosage Administration Record (DAR) eCRF are considered randomized treatment. ***Crossover treatment*** will refer to the study treatment received during the crossover treatment period.

Date of first administration of randomized treatment

The date of first administration of randomized treatment is derived as the first date when a nonzero dose of randomized treatment was administered as per the DAR eCRF. The date of first administration of randomized treatment will also be referred as ***start of randomized treatment***.

Date of first administration of crossover treatment

The date of first administration of crossover treatment is derived as the first date when a nonzero dose of crossover treatment was administered as per the DAR eCRF. The date of first administration of crossover treatment (ruxolitinib) will also be referred as ***start of crossover treatment***.

Date of last administration of randomized treatment

The date of last administration of randomized treatment is defined as the last date when a nonzero dose of randomized treatment was administered as per DAR eCRF.

Date of last administration of crossover treatment

The date of last administration of crossover treatment is defined as the last date when a nonzero dose of crossover treatment (ruxolitinib) was administered as per DAR eCRF.

Study day

The study day, describes the day of the event or assessment date, relative to the reference start date.

The study day is defined as:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

The reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, PK, etc.), [redacted] and PRO is the start of randomized treatment.

The reference start date for all other, non-safety assessments (i.e., aGvHD assessment, survival, aGvHD progression, aGvHD response, underlying hematologic disease relapse/progression, etc.) is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Crossover study day

The crossover study day, describes the day of the event or assessment date, relative to the start of crossover treatment.

Crossover study day = date of event – start of crossover treatment + 1, if event is on or after the start of crossover treatment

Crossover study day = date of event – start of crossover treatment, if event precedes the start of crossover treatment

The crossover study day will be displayed in the data listings if an event starts on or after the start of crossover treatment.

Time unit

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

Baseline

Baseline for safety endpoints (except adverse events), [redacted] and patient reported outcomes is defined as the last assessment prior to or on the treatment start date.

For evaluations after cross over, the baseline is defined as the last assessment prior to or on the start date of crossover treatment.

Baseline for efficacy endpoints is defined as the last assessment or procedure conducted prior to or on the date of randomization date + 3 days, but no later than the treatment start date.

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

On-treatment assessment/event and observation periods

For adverse event reporting the overall observation period will be divided into four mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent at screening to the day before first administration of study treatment
2. ***on-randomized-treatment period***: from date of first administration of randomized treatment to 30 days after date of last actual administration of randomized treatment (including start and stop date) or end of randomized treatment per End of Randomized Treatment Disposition eCRF, whichever is later; for those patients who cross over from BAT to ruxolitinib, the period is from date of first administration of randomized treatment to earlier of (i) 30 days after date of last actual administration of randomized treatment or end of randomized treatment per End of Randomized Treatment Disposition eCRF, whichever is later, (ii) the day before the date of first administration of crossover treatment. Up to Day 28 visit, more than one BAT regimen may be initiated as study treatment. In this case, the last actual administration of randomized treatment refers to the last actual administration of the last BAT regimen reported on the DAR eCRF.

3. ***on-crossover-treatment period***: from date of first administration of crossover treatment to 30 days after date of last administration of crossover treatment (including start and stop date) or end of crossover treatment per End of Crossover Treatment Disposition eCRF, whichever is later.
4. ***post-treatment period***: starting at Day 31 after last administration of study treatment or the day after end of study treatment per end of treatment disposition eCRFs, whichever is later.

If dates are incomplete in a way that clear assignment to pre-, on-, post-treatment period cannot be made, then the respective data will be assigned to the on-randomized-treatment period.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). The data on-randomized-treatment period and on-crossover-treatment period will be summarized separately. In addition, a separate summary for death including on-treatment and post-treatment deaths will be provided. In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period (***treatment-emergent*** AEs).

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

Windows for multiple assessments

In order to summarize data collected over time (including unscheduled visits), the assessments will be time slotted. The following general rule will be applied in creating the assessment windows: If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If two assessments within a time window are equidistant from the target date, then the later of the two assessments will be used. If multiple assessments on the same date then the worst case will be used. Data from all assessments (scheduled and unscheduled), including multiple assessments, will be listed.

The following time windows are defined for descriptive summary on aGvHD assessment, [redacted] PROs and safety (Table 2-1) by visit. The end of treatment assessment will be mapped into the time points if collected within 7 days of the last dose intake.

Table 2-1 Time windows for aGvHD assessment, [redacted] PROs and safety assessment (lab, vital sign, etc.)

Time Window	Planned Visit Timing	Time Window Definition
On treatment		
Baseline[1]	Study Day 1[1]	≤ Study Days 1+3, no later than treatment start date
Baseline[2]	On or before Study Day 1[2]	≤ Study Day 1
Week 1	Study Day 7	Study Days 4 – 10
Week 2	Study Day 14	Study Days 11 – 17
Week 3	Study Day 21	Study Days 18 – 24
Week 4	Study Day 28	Study Days 25 – 31
Week 5	Study Day 35	Study Days 32 – 38

Week 6	Study Day 42	Study Days 39 – 45
Week 7	Study Day 49	Study Days 46 – 52
Week 8	Study Day 56	Study Days 53 – 59
Week 12	Study Day 84	Study Days 71 – 98
Week 16	Study Day 112	Study Days 99 – 126
Week 20	Study Day 140	Study Days 127 – 154
Week 24	Study Day 168	Study Days 155 – 182
Safety follow-up	30 days after last dose	Last dose date + 30

Baseline[1] for aGvHD assessment (efficacy);
Baseline[2] for PROs, safety assessments [redacted] (safety);
Study Day 1[1] = randomization date;
Study Day 1[2] = start date of randomized treatment;
Crossover Study Day 1 = start date of crossover treatment
EOT (randomized treatment or crossover treatment) assessments are mapped to the time points.
Safety follow-up is a separate time point for PROs and safety assessments.

Last contact date

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

Table 2-2 Last contact date data sources

Source data	Conditions
Date of Randomization	No Condition
Last date patient was known to be alive from Survival Follow-up page	Patient status is reported to be alive, lost to follow-up or unknown.
Start/End dates from further aGvHD therapy	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
End of treatment date from end of treatment page	No condition.
- aGvHD assessment date - any specific efficacy assessment date if available (e.g., cGvHD assessment, graft failure assessment, hematologic disease relapse/progression assessment)	Evaluation is marked as ‘done’.
Laboratory/PK collection dates/[redacted]	Sample collection marked as ‘done’.
Vital signs date	At least one non-missing parameter value
Start/End dates of AE	Non-missing verbatim term

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date . The cut-off date will not be used for last contact date, unless the patient was seen or contacted on that date. No date post cut-off date will be used. Completely 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. Partial date imputation is allowed for event (death)/censoring coming from 'Survival information' eCRF.

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

2.2 Analysis sets

Full Analysis Set

The Full Analysis Set (FAS) comprises all patients to whom study treatment has been assigned by randomization. According to the intent to treat principle, patients will be analyzed according to the treatment and strata, they have been assigned to during the randomization procedure.

Per-Protocol Set

The Per-Protocol Set (PPS) consists of a subset of the patients in the FAS who are compliant with requirements of the clinical study protocol.

The following list of protocol deviations will lead to exclusion of the patient from the Per-Protocol Set:

- Not steroid refractory aGvHD (PDID: INCL09 per SSD v22.0)
- More than one prior systemic therapy for the treatment of aGvHD other than corticosteroids +/- CNI (prophylaxis or treatment) (PDID: EXCL01 per SSD v22.0)
- Missing or incorrect aGvHD grade at randomization (PDID: OTH02, INCL07 per SSD v22.0)
- Taking any prohibited medication as specified in this protocol after start of study treatment and before end of study treatment (PDID: COMD01, COMD02 per SSD v22.0)
- Study treatment received different from treatment arm assigned by randomization (PDID: TRT02 per SSD v22.0)

Safety Set

The Safety Set includes all patients who received at least one dose of study treatment. Patients will be analyzed according to the study treatment received, where treatment received is defined as the randomized treatment if the patient took at least one dose of that treatment or the first treatment received if the randomized treatment was never received during the randomized treatment period.

Crossover Analysis Set

The Crossover Analysis Set (CAS) comprises all patients randomized to and received BAT who then crossed over and received at least one dose of ruxolitinib. This analysis set will be used for all analyses for crossover patients.

Pharmacokinetic Analysis Set

The Pharmacokinetic Analysis Set (PAS) includes all patients who provide at least one evaluable PK concentration. For a concentration to be evaluable, patients are required to:

- Take a dose of ruxolitinib prior to sampling.
- For pre-dose samples, do not vomit within 2 hours after the dosing of ruxolitinib prior to sampling; for post-dose samples, do not vomit within 2 hours after the dosing of ruxolitinib.

The PAS will be used for NCA analysis for patients where extensive PK sampling is obtained [redacted]

Patient Classification:

Patients may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific patient classification rules defined in Table 2-3.

Table 2-3 Patient classification based on protocol deviations and non-PD criteria

Analysis set	Protocol deviations leading to exclusion	Non protocol deviation leading to exclusion
FAS	INCL03	Not applicable
Safety Set	INCL03	No dose of study treatment
Per-Protocol Set	INCL03, INCL07, INCL09, EXCL01, TRT02, , OTH02, During randomized treatment period: COMD01, COMD02	No dose of study treatment
Crossover Analysis Set	INCL03	No dose of ruxolitinib
PK Analysis Set	INCL03	No dose of ruxolitinib, No evaluable PK concentration

Note: Based on CINC424C2301_Study Specification Document version 22.0.
INCL03 - Written Study informed consent /assent not obtained.

Withdrawal of Informed Consent

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

2.2.1 Subgroup of interest

Efficacy

The primary efficacy endpoint will be summarized by the following subgroups to examine the homogeneity of treatment effect provided that the primary efficacy analysis based on the FAS is statistically significant :

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race
- Region (Europe, Australia and Canada), Japan, Asia excluding Japan)
- Acute GvHD grade (Grade II, III, IV)
- Source of grafts (related, not related)
- Criteria for SR-aGvHD (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper)
- Prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/- other systemic aGvHD treatment,)
- Conditioning regimen type (myeloablative, non-myeloablative, reduced intensity)
- Stem cell type (bone marrow, peripheral blood, single cord blood)
- Donor HLA status (match, mismatch)
  - Note: HLA status Match score means that 1:1 match, e.g. 10/10, 8/8, 6/6, 12/12. Mismatch will be all the other scores.
- Donor gender match (F/F, M/M, F/M, M/F)
- Donor CMV status (negative, positive)
- Donor source/HLA match status (related and matched, related and mismatched, not related and matched, not related and mismatched)
- aGvHD organ involvement at randomization (skin, liver, upper GI, lower GI)

No formal statistical test of hypotheses will be performed for the subgroups, only point estimate of the treatment effect and 95% confidence intervals will be provided (see Section 2.5 for further analysis details). The objective of the efficacy subgroup analysis is to demonstrate homogeneity of treatment effect in the above subgroups.

Safety

Key safety analyses will be repeated on the Safety Set in the following subgroups:

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race
- Region (Europe, Australia and Canada), Asia excluding Japan, Japan)
- aGvHD organ involvement at randomization (skin, liver, upper GI, lower GI)

The objective for carrying out these subgroup analyses is to identify potential safety issues that may be limited to or more commonly observed in a subgroup of patients. The following summaries will be presented by subgroup:

- AEs, irrespective of causality, by primary system organ class and preferred term
- AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- Serious AEs, irrespective of causality, by primary system organ class and preferred term
- Serious AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- On-treatment deaths, by primary system organ class and preferred term

Adolescent patients

In order to be able to make a separate risk/benefit assessment for the adolescent patients, besides the above planned subgroup analyses for this age group (12-<18 years), data of demographics and exposure will be presented.

Japanese patients

Subgroup analyses will also be performed for the patients treated in Japan. No selection will be done on the basis of ethnicity, the purpose being to evaluate the population of patients living in Japan, not a specific ethnic set of patients. The analysis will be done the same way as in region subgroup specified above.

2.3 Patient disposition, demographics and other baseline characteristics

The Full Analysis Set (FAS) will be used for all baseline and demographic summaries and listings unless otherwise specified. Summaries will be reported by treatment arm and for all patients and listings will be reported by treatment arm to assess baseline comparability. No inferential statistics will be provided.

Basic demographic and background data

All demographic and baseline disease characteristics data will be summarized and listed by treatment arm. Categorical data (e.g. gender, age groups: 12- <18 vs. 18-65 vs. >65 years, race, ethnicity) will be summarized by frequency counts and percentages; the number and percentage of patients with missing data will be provided. Continuous data (e.g. age, weight, height, body mass index) will be summarized by descriptive statistics (N, mean, median, standard deviation, Q1, Q3, minimum and maximum).

Baseline stratification factors

The number (%) of patients in each stratum (Grades II, III, IV) based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS. Discordances between the stratum recorded in IRT at the time of randomization and the actual stratum recorded in the clinical database through the data collected on eCRF will be cross-tabulated and listed.

Diagnosis and extent of disease

Summary statistics will be tabulated for diagnosis and extend of disease in underlying disease, stem cell transplant and acute GvHD.

For underlying disease, the analysis will include the following: primary diagnosis category and subcategory, details of primary diagnosis, time since diagnosis of underlying disease, CIBMTR risk assessment.

For transplant related disease history, the analysis will include the following: conditioning regimen type, total HCT-specific comorbidity index score, time since transplant, time from diagnosis of underlying disease to transplant, stem cell type, cytomegalovirus status, donor information including age, gender, HLA typing method, HLA match score, source of grafts (related/unrelated), CMV status, T-cell depleted (Y/N), total nucleated cell dose.

For aGvHD disease history, the analysis will include the following: time since diagnosis of aGvHD grade ≥2, aGvHD grade when diagnosis of grade ≥2, steroid refractory aGvHD criteria met (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper), prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/-other systemic aGvHD treatment,), time from diagnosis of aGvHD grade ≥2 to steroid refractory, time since steroid refractory aGvHD, aGvHD grade at randomization, aGvHD organ involvement, steroid dose at randomization.

Medical history

Medical history and ongoing conditions, including underlying disease conditions and symptoms entered on eCRF will be summarized and listed by treatment arm. The summaries will be presented by primary system organ class (SOC), preferred term (PT) and treatment arm. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

Other

Intended BAT strategy in case of SR aGvHD prior to randomization will be summarized by treatment arm.

All data collected at baseline including child bearing potential will be listed.

2.3.1 Patient disposition

Enrollment by country and center will be summarized for all screened patients and also by treatment arm. The number (%) of randomized patients will be presented overall and by treatment group. The number (%) of screened and not-randomized patients and the reasons for screening failure will also be displayed. The number (%) of patients in the FAS who are still on treatment, who discontinued the treatment phases (randomized treatment, crossover treatment) as well as the reason for discontinuation, and the survival follow-up will be presented overall and by treatment group.

Protocol deviations

The number (%) of patients in the FAS with any protocol deviation will be tabulated by deviation category (as specified in the study Data Handling Plan) overall and by treatment group for the FAS. All protocol deviations will be listed.

Analysis sets

The number (%) of patients in each analysis set (defined in Section 2.2) will be summarized by treatment group and stratum.

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

2.4.1 Study treatment / compliance

Duration of exposure in days to ruxolitinib and BAT will be summarized by means of descriptive statistics. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. Duration of exposure to each BAT regimen (anti-thymocyte globulin (ATG), extracorporeal photopheresis (ECP), mesenchymal stromal cells (MSC), low-dose methotrexate (MTX), mycophenolate mofetil (MMF), mTOR inhibitors (everolimus or sirolimus), etanercept, or infliximab) will be summarized using the same approach.

Actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized for the ruxolitinib arm. Patients randomized to Investigator's choice of BAT will receive various different categories of therapy.

The number (%) of patients who have dose changes or interruptions, and the reasons, will be summarized for ruxolitinib group.

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

The Safety Set and Crossover Analysis Set will be used for all summaries on randomized treatment and crossover treatment, respectively. The Safety Set will be used for listings of study treatment.

Duration of exposure to study treatment

Duration of exposure to randomized treatment (days) = (last date of exposure to randomized treatment) – (date of first administration of randomized treatment) + 1.

Duration of exposure to crossover treatment (days) = (last date of exposure to crossover treatment) – (date of first administration of crossover treatment) + 1.

The last date of exposure to study treatment is the latest of the last dates of exposure to study treatment (see Table 2-4).

Table 2-4 Definition of last date of exposure of study treatment

Scenario	Definition of last date of exposure of study treatment	Example
Scenario 1: Study treatment with a periodical administration	The planned end date of the last period in which the last non-zero dose of the study treatment was last administered.	Example 1: In a once-a-week administration, the last date of exposure is the date of last administration + 6 days.

	Note : If the patient changed to another BAT, died or was lost to follow-up before the derived last date, the last date of exposure to study treatment is the date prior to the start of the next BAT, date of death or the date of last contact, respectively. If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	Example 2: In a twice-a-week administration, the last date of exposure is the date of last administration + 3 days.
Scenario 2: Study treatment with daily/IV administration	Date of last administration of a non - zero dose of the study treatment.	Example 3: A patient had a permanent discontinuation of the study treatment 06Jan2017 after being put on a temporary interruption since 01Jan2017. In this case the last date of exposure is- 31Dec2016.
Scenario 3: Study treatment as an antibody	Date of last administration of a non - zero dose of the study drug + number of days antibody persists in vivo – 1 day. Note : If the patient changed to another BAT, died or was lost to follow-up before the derived last date, the last date of exposure to study treatment is the date prior to the start of the next BAT, date of death or the date of last contact, respectively. If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	Example 4: For a study treatment which antibody persists in vivo for 28 days, the last date of exposure is the date of last administration + 28 days – 1 day.

Summary of duration of exposure of study treatment in days will include categorical summaries (based on clinically meaningful time intervals) and continuous summaries (i.e. mean, standard deviation etc.) using appropriate units of time.

Duration of exposure in patient-years

The duration of exposure in patient-years is a total of the duration of exposure in years from all the patients in a treatment group. It will be calculated for randomized treatment (by treatment group) and crossover treatment, respectively.

Duration of treatment period

Duration of randomized treatment period (days) = end date of on-randomized-treatment period – date of first administration of randomized treatment + 1

Duration of crossover treatment period (days) = end date of on-crossover-treatment period – date of first administration of crossover treatment + 1

The on-randomized-treatment period and on-crossover-treatment period are defined in Section 2.1.1.

The duration of randomized treatment period in days for ruxolitinib and BAT will be summarized by means of descriptive statistics. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The duration of crossover treatment period will be summarized similarly. As patients randomized to BAT may crossover between Day 28 and Week 24, the time from the randomization to the first dose of crossover treatment will also be summarized.

Duration of treatment period in patient-years

The duration of treatment period (randomized and crossover) in patient-years is a total of the duration of treatment period in years from all the patients in a treatment group. It will be calculated for randomized treatment (ruxolitinib vs. BAT) and crossover treatment separately.

Cumulative dose

The **planned cumulative dose** for ruxolitinib refers to the total planned dose as per the protocol (10 mg bid) up to the last dose date.

The **actual cumulative dose of randomized ruxolitinib** refers to the total actual dose of randomized ruxolitinib as documented in the DAR eCRF.

The **actual cumulative dose of crossover ruxolitinib** refers to the total actual dose administered, over the duration for which the patient is on the crossover ruxolitinib as documented in the DAR eCRF.

For patients who did not take any drug the cumulative dose is by definition equal to zero.

Dose intensity and relative dose intensity of ruxolitinib

Dose intensity (DI) **of randomized ruxolitinib** for patients with non-zero duration of exposure is defined as follows:

DI (mg / day) = Actual cumulative dose (mg) of randomized ruxolitinib / Duration of exposure to randomized ruxolitinib (days).

Dose intensity (DI) **of crossover ruxolitinib** for patients with non-zero duration of exposure is defined as follows:

DI (mg / day) = Actual cumulative dose (mg) of crossover ruxolitinib / Duration of exposure to crossover ruxolitinib (days).

For patients who did not take any drug the DI is by definition equal to zero.

Planned dose intensity (PDI) is defined as follows:

PDI (mg / day) = Planned Cumulative dose (mg) / Duration of exposure (days).

The protocol planned starting dose for ruxolitinib is 10 mg BID.

Relative dose intensity (RDI) is defined as follows:

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

DI and RDI will be summarized for randomized and crossover ruxolitinib treatment, separately.

The actual cumulative dose, DI and RDI up to Day 28 visit, Day 56 visit and last date of exposure to study treatment (randomized or crossover) will be summarized.

The number (%) of patients at total daily dose 5 mg, 10 mg, 15 mg and 20 mg will be summarized at Day 28 visit, Day 56 visit and last date of exposure to study treatment.

Dose reductions, interruptions or permanent discontinuations

The number of patients who have dose changes, or interruptions, and the reasons, will be summarized for ruxolitinib (randomized and crossover treatment separately). The number of patients who have dose permanent discontinuations and the reasons, will be summarized by treatment group.

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

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

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

2.4.2 Prior, concomitant and post therapies

2.4.2.1 Prior aGvHD treatment

The number and percentage of patients who received any prior aGvHD treatment (medications and procedures) will be summarized by lowest ATC class, preferred term and treatment arm.

Listings will be produced for prior aGvHD treatment.

The above analyses will be performed using the FAS.

2.4.2.2 Prior prophylaxis

The number and percentage of patients who received any prophylaxis prior to randomization will be summarized by lowest ATC class, preferred term and treatment arm using FAS.

Listings will be generated for prophylaxis.

2.4.2.3 Systemic corticosteroid

The duration of exposure and the peak dose will be summarized for pre-randomizatition period, between screening and randomization.

The duration of exposure will be summarized for on-randomized-treatment period and on-crossover-treatment period separately. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The actual cumulative dose, dose intensity and relative dose intensity (relative to the starting dose of corticosteroids) will be summarized up to Day 28 visit, Day 56 visit and end of on-treatment period. These analyses will be based on Safety Set. The data will be reported on the DAR eCRF through the end of randomized treatment per EOT Disposition eCRF, and on the Prior and Concomitant Medications eCRF after the end of randomized treatment. The list of systemic corticosteroid will be documented in Data Handling Plan.

2.4.2.4 Calcineurin inhibitors (CNIs) during study treatment

The duration of exposure will be summarized for CNIs (cyclosporine or tacrolimus) during on-randomized-treatment period and on-crossover-treatment period, respectively. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. These analyses will be based on Safety Set. The data will be reported on the DAR eCRF through the end of randomized treatment per EOT Disposition eCRF, and on the Prior and Concomitant Medications eCRF after the end of randomized treatment. The list of CNIs will be documented in Data Handling Plan.

2.4.2.5 Additional systemic aGvHD therapy

New additional systemic aGvHD therapy (medications and procedures) since start of study treatment will be listed and summarized by lowest ATC class, preferred term, overall and by treatment group by means of frequency counts and percentages using FAS.

2.4.2.6 Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) other than the study treatment administered to a patient coinciding with the study treatment period. Concomitant therapy include medications (other than study drugs) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system and summarized by lowest ATC class and preferred term using frequency counts and percentages. Surgical and medical procedures will be coded using MedDRA and summarized by SOC and preferred term. The summaries for randomized treatment phase using Safety Set will include:

- Medications starting on or after the start of randomized treatment but no later than end of on-randomized-treatment period; and
- Medications starting prior to start of randomized treatment and continuing after the start of randomized treatment.

The summaries for crossover treatment phase using Crossover Analysis Set will include:

- Medications starting on or after the start of crossover treatment but no later than end of on-crossover-treatment period; and
- Medications starting prior to start of crossover treatment and continuing after the start of crossover treatment.

All concomitant therapies will be listed using Safety Set. Any concomitant therapies starting and ending prior to the start of randomized treatment or starting beyond end of on-randomized-treatment period if not crossed over, or starting beyond end of on-crossover-treatment period if crossed over, will be flagged in the listing.

The prohibited concomitant medications will be summarized by lowest ATC class and preferred term up to the end of on-randomized-treatment and on-crossover-treatment periods, respectively. The list of prohibited medications will be provided and updated regularly by clinical team according to the clinical database review. The topical medications are excluded from the list.

In addition, a subset of concomitant medications i.e. transfusions (red blood cells and platelets) will be grouped and summarized by treatment group.

2.5 Analysis of the primary objective

The primary objective of the study is to compare the overall response rate (ORR) at Day 28 between the ruxolitinib arm and BAT arm in steroid refractory aGvHD patients.

2.5.1 Primary endpoint

The primary endpoint was reported in the primary analysis, and will not be repeated in this second analysis, except for the subgroup analyses.

ORR at Day 28 is defined as the proportion of patients with complete response (CR) or partial response (PR) at Day 28 according to standard criteria [Harris 2016]. Note that response is relative to the assessment of aGvHD at randomization.

- **Complete response** is defined as a score of 0 for the aGvHD grading in all evaluable organs that indicates complete resolution of all signs and symptoms of aGvHD in all evaluable organs without administration of additional systemic therapies for any earlier progression, mixed response or non-response of aGvHD.
- **Partial response** is defined as improvement of 1 stage in 1 or more organs involved with aGvHD signs or symptoms without progression in other organs or sites without administration of additional systemic therapies for an earlier progression, mixed response or non-response of aGvHD.
- **Lack of response** is defined as no response, mixed response, or progression.
    - **No response** is defined as absence of improvement in any organ involved by aGvHD, without worsening in any involved organ.
    - **Mixed response** is defined as improvement of at least 1 stage in the severity of aGvHD in at least one organ accompanied by progression in another organ or development of signs or symptoms of aGvHD in a new organ.

- **Progression** is defined as worsening in 1 or more organs by 1 or more stages without improvement in any involved organ

A flare in aGvHD is defined as any increase in signs or symptoms of aGvHD after an initial response (CR or PR). However, a flare may not lead to progression or additional systemic therapy. Only flares in GvHD that require new additional systemic therapy, will be considered aGvHD flare failure. Patients who fail corticosteroid taper fulfilling either one of the following criteria should initiate additional systemic therapy:

- Requirement for an increase in the corticosteroid dose to methylprednisolone ≥2 mg/kg/day (or equivalent prednisone dose ≥2.5 mg/kg/day), OR
- Failure to taper the methylprednisolone dose to <0.5 mg/kg/day (or equivalent prednisone dose <0.6 mg/kg/day) for a minimum 7 days.

A patient will not be considered a responder at Day 28 if any of the following events occurs:

- Missing aGvHD assessment at baseline or Day 28
- No CR or PR at Day 28
- Additional systemic therapy for aGvHD prior to Day 28

Compared to the protocol version 00, 'discontinuation from randomized treatment' has been dropped as reason. Since some BATs have fixed duration which is less than 8 weeks, to avoid bias, discontinuation from randomized treatment will not be used to disqualify the response or durable response. Also duration of response will not be censored based on treatment discontinuation.

Acute GvHD will be assessed according to standard criteria [Harris 2016], as described in protocol Appendix 1. Grade will be calculated based on the staging of the organs and recorded on the eCRF by the Investigator. The Investigator reported grade will be used for randomization and reported grade and response for all analyses. Grade and response will be calculated by the sponsor for the purposes of data review only and sensitivity analysis.

2.5.2 Statistical hypothesis, model, and method of analysis

The primary efficacy analysis is the comparison of ORR at Day 28 between the two treatment arms. The following statistical hypotheses will be tested to address the primary efficacy objective:

H_0: $ORR_{rux} \leq ORR_{BAT}$ vs. H_1: $ORR_{rux} > ORR_{BAT}$

where ORR_{rux} and ORR_{BAT} are the overall response rates at Day 28 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., aGvHD Grade II vs III vs. IV), will be used to compare ORR between the two treatment groups, at the one-sided 2.5% level of significance.

The primary efficacy variable, ORR at Day 28, will be analyzed at the time when all patients have completed their Day 56 visit or discontinued earlier. The primary analysis will be performed on FAS according to ITT principle. ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. One-sided p-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

2.5.3 Handling of missing values/censoring/discontinuations

Patients with missing assessments that prevent the evaluation of the primary endpoint will be considered non-responders on that treatment arm. This includes missing aGvHD response assessments at baseline and Days 28, 56.

The following analysis windows (also in Table 2-1) will be applied to the target day for assessments on overall response, where target day for Week X is X*7.

Baseline assessment is the last aGvHD assessment prior to or on the date of randomization (Day 1). A tolerance of up to 3 days from randomization will be considered in the analysis, but the baseline assessment will not be later than the date of treatment start.

Weeks 1, 2, 3, 4, 5, 6, 7, 8: -3 days/+3 days

Weeks 12 to 24: -13 days/+14 days

The analysis windows for assessments after crossover is similar, except that the baseline is the last aGvHD assessment prior to or on Crossover Day 1 (date of first administration of crossover treatment).

No data imputation will be applied.

2.5.4 Supportive analyses for ORR at Day 28

[redacted]

For each of the subgroups, the following analyses will be performed:

- Proportion of patients with ORR using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]
- Odds ratio with 95% CI using a logistic regression model with treatment and stratification factors as covariate

Efficacy analyses in subgroups will be purely exploratory and are intended to explore the consistency of treatment effect. Forest plot (n, odds ratio, 95% CI) will be produced to graphically depict the treatment effect estimates in different subgroups. No inferential statistics (p-values) will be produced for the subgroups.

[redacted]

2.5.5 ORR at Crossover Day 28

ORR at Crossover Day 28 is defined as the proportion of crossover patients with complete response (CR) or partial response (PR) at Crossover Day 28 according to standard criteria [Harris 2016]. ORR will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] based on the CAS using local investigators' review of aGvHD assessment data. Note that response is relative to the last assessment of aGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a responder at Crossover Day 28 if any of the following events occurs:

- Missing aGvHD assessment at Crossover baseline or Crossover Day 28
- No CR or PR at Crossover Day 28
- Additional systemic therapy for aGvHD prior to Crossover Day 28

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Crossover baseline to Crossover Day 28.

2.6 Analysis of the key secondary objective

The key secondary objective of the study is to determine whether treatment with ruxolitinib has better durable ORR at Day 56 compared with BAT.

2.6.1 Key secondary endpoint

The key secondary endpoint was reported in the primary analysis, and will not be repeated in this second analysis.

Durable ORR at Day 56 is defined as the proportion of all patients in each arm who achieve a complete response (CR) or partial response (PR) at Day 28 and maintain a CR or PR at Day 56. If a patient is a CR at Day 28 and a PR at Day 56, he/she will be considered as a durable responder. A patient will not be considered a durable responder at Day 56 if any of the following events occurs:

- Not a responder at Day 28
- Missing aGvHD assessment at Day 56
- No CR or PR at Day 56.
- Additional systemic therapy for aGvHD prior to Day 56

Durable ORR will be calculated based on the FAS using local investigators review of aGvHD assessment data.

The patients randomized to BAT who meet cross-over criteria and cross-over to ruxolitinib are considered to have the additional systemic therarpy for aGvHD, and will not be considered as a responder afterwards.

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Day 28 to Day 56 value for the patients who achieved PRs at Day 28.

2.6.2 Statistical hypothesis, model, and method of analysis

The following statistical hypotheses will be tested to address the key secondary efficacy objective:

H_0: $DORR_{rux} \leq DORR_{BAT}$ vs. H_1: $DORR_{rux} > DORR_{BAT}$

where $DORR_{rux}$ and $DORR_{BAT}$ are the durable overall response rates at Day 56 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., aGvHD Grade II vs III vs. IV), will be used to compare the durable ORR between the two treatment groups, at the one-sided 2.5% level of significance, if the primary endpoint is significant.

The key secondary efficacy variable, durable ORR at Day 56, will be analyzed at the time when all patients have completed their Day 56 visit or discontinued earlier. The analysis will be based on FAS according to ITT principle. Durable ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. P-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

The durable ORR at Day 56 will be tested hierarchically. That is, if the ORR at Day 28 is statistically significant, the durable ORR at Day 56 will be tested. If the ORR at Day 28 is not statistically significant, the durable ORR at Day 56 will not be tested.

2.6.3 Handling of missing values/censoring/discontinuations

Refer to Section 2.5.3

2.6.4 Durable ORR at Crossover Day 56

Durable ORR at Crossover Day 56 is defined as the proportion of all crossover patients who achieve a complete response (CR) or partial response (PR) at Crossover Day 28 and maintain a CR or PR at Crossover Day 56. Durable ORR at Crossover Day 56 will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] based on the CAS using local investigators' review of aGvHD assessment data. Note that response is relative to the last assessment of aGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a durable responder at Crossover Day 56 if any of the following events occurs:

- Not a responder at Crossover Day 28
- Missing aGvHD assessment at Crossover Day 56
- No CR or PR at Crossover Day 56.
- Additional systemic therapy for aGvHD prior to Crossover Day 56

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Crossover Day 28 to Crossover Day 56 value for the patients who achieved PRs at Crossover Day 28.

2.7 Analysis of secondary efficacy objective(s)

The other secondary efficacy objectives are to:

- Evaluate the two treatment arms with respect to ORR at other time points, e.g. Day 14
- Evaluate the two treatment arms with respect to duration of response (DOR)
- Evaluate the two treatment arms with respect to overall survival (OS)
- Evaluate the two treatment arms with respect to event free survival (EFS)
- Evaluate the two treatment arms with respect to failure free survival (FFS)
- Evaluate the two treatment arms with respect to Non-relapse mortality (NRM)
- Evaluate the two treatment arms with respect to incidence of malignancy relapse/progression (MR)
- Describe cumulative steroid dosing until Day 56 in each treatment arm
- Evaluate the two treatment arms with respect to incidence of cGvHD
- Evaluate the two treatment arms with respect to BOR at any time points up to Day 28

All the secondary efficacy endpoint analyses are non-comparative in nature and will be analyzed using the Full Analysis Set (FAS).

2.7.1 Secondary efficacy endpoints

Overall Response Rate at Day 14

Overall Response Rate at Day 14 is defined as the proportion of patients with complete response (CR) or partial response (PR) at Day 14 according to standard criteria [Harris 2016]. This endpoint was reported in the primary analysis, and will not be repeated in this second analysis.

Duration of response (DOR)

Duration of response is defined for patients whose overall response at Day 28 is complete response (CR) or partial response (PR) according to standard criteria [Harris 2016]. The start date is the date of first documented response of CR or PR (i.e., the start date of response), which could be prior to or at Day 28. If it's prior to Day 28, there should not be progression or addition of systemic therapies for aGvHD between the start date of response and Day 28. The end date is defined as the date of progression or the date of addition of systemic therapies for aGvHD on or after Day 28.

Death without prior observation of aGvHD progression and onset of chronic GvHD are considered to be competing risks.

Duration of response will be censored at the last response assessment prior to or at the analysis cut-off date, if no events/competing risk occurred before or at the cut-off date.

Overall survival (OS)

Overall survival is defined as the time from date of randomization to date of death due to any cause. If a patient is not known to have died, then OS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Event Free Survival (EFS)

Event-free survival is defined as the time from the date of randomization to the date of hematologic disease relapse/progression, graft failure or death due to any cause. If a patient is not known to have any event, then EFS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Failure Free Survival (FFS)

Failure-free survival is defined as the time from the date of randomization to date of hematologic disease relapse/progression, non-relapse mortality or addition of new systemic aGvHD treatment.

Non-relapse mortality (NRM)

Non-relapse mortality is defined as the time from date of randomization to date of death not preceded by hematologic disease relapse/progression. Hematologic disease relapse/progression is considered a competing risk for NRM with the date of hematologic disease relapse/progression being the earlier of documented hematologic disease relapse/progression or institution of therapy to treat potential hematologic disease relapse/progression. If a patient is not known to have died or to have relapsed/progressed, then NRM will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Incidence of Malignancy Relapse/Progression (MR)

Malignancy relapse/progression is defined as the time from date of randomization to hematologic malignancy relapse/progression. Deaths not preceded by hematologic malignancy relapse/progression are competing risks. If a patient is not known to have event or competing risks, then MR will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Incidence of cGvHD

cGvHD is defined as the diagnosis of any cGvHD including mild, moderate, severe. Incidence of cGvHD is the time from date of randomization to onset of cGvHD. Deaths without prior onset of cGvHD and hematologic disease relapse/progression are competing risks. If a patient is not known to have event or competing risks, then the incidence of cGvHD will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Best Overall Response

Best overall response (BOR) is defined as proportion of patients who achieved overall response (CR or PR) at any time point up to and including Day 28, and had no additional systemic therapy for aGvHD prior to the time point. This endpoint was reported in the primary analysis, and will not be repeated in this second analysis.

2.7.2 Statistical hypothesis, model, and method of analysis

Overall Response Rate at Day 14

Overall response rate will be calculated based on the local investigators review of aGvHD assessment data. It will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

Duration of response (DOR)

The estimated cumulative incidence rates and 95% confidence intervals at 1, 2, 6, 12, 18 and 24 months will be presented for each treatment group. The cumulative incidence curve will be plotted. DOR will be listed and summarized by treatment group for all patients in the FAS with overall response of CR or PR at Day 28.

Overall survival (OS)

OS will be analyzed according to the randomized treatment group and strata assigned at randomization (aGvHD grade: Grade II vs. III vs. IV). The OS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 1, 2, 6, 12, 18 and 24 month survival estimates and 95% confidence intervals [Brookmeyer and Crowley 1982] will be presented for each treatment group. The hazard ratio for OS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

OS follow-up time will be summarized by treatment group for all patients in FAS. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval.

Event Free Survival (EFS)

EFS will be analyzed according to the randomized treatment group and strata assigned at randomization. The EFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 1, 2, 6, 12, 18 and 24 month survival estimates and 95% confidence intervals [Brookmeyer and Crowley 1982] will be presented for each treatment group. The hazard ratio for EFS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

A sensitivity analysis including aGvHD progression as an event will be performed using the same approach.

Failure Free Survival (FFS)

Cumulative incidence curve for FFS as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for each treatment group, accounting for onset of chronic GvHD as the competing risk.

The FFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curveswill be presented for each treatment group. The hazard ratio for FFS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

In addition, the cumulative incidence of each of the three components considering the other two components as a competing risks will be estimated. Onset of chronic GvHD is considered as a competing risk for all three types of failure. The cumulative incidence curves will be plotted for each treatment group.

A sensitivity analysis including aGvHD progression as an event will be performed using the same approach.

Non-relapse mortality (NRM)

NRM will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization (aGvHD grade: Grade II vs. III vs. IV). The cumulative incidence curve for NRM as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for each treatment group.

As a sensitivity analysis, the cumulative incidence curve for NRM as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for patients with underlying hematologic malignant disease in each treatment group.

Incidence of Malignancy Relapse/Progression (MR)

The cumulative incidence curve for MR and estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for patients with underlying hematologic malignant disease, accounting for NRM as the competing risk.

In addition, the proportion of patients who had hematologic malignancy relapse/progression and its 95% confidence interval at 1, 2, 6, 12, 18 and 24 months will be presented by treatment group for patients with underlying hematologic malignant disease. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

Cumulative steroid dosing until Day 56

Overall and weekly cumulative steroid dose for each patient up to Day 56 or discontinuation of randomized treatment will be tabulated. In addition, the relative dose intensity (RDI), by week, will be calculated relative to the starting dose of corticosteroids and categorized as (1) complete reduction where patients are tapered off corticosteroids by D56, (2) ≤50% RDI and (3) >50% RDI. The proportion of patients in each category and corresponding 95% confidence intervals will be presented by treatment group. Odds ratio and 95% confidence limits calculated from Fisher's exact test will be also presented for the proportion of patients who are completely tapered off corticosteroids by D56. Average corticosteroid dose during the week ending on Days 14, 28, 56, 84, and 168 will also be tabulated and plotted.

Incidence of cGvHD

The cumulative incidence of cGvHD and estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented, accounting for competing risks. The cumulative incidence curve will be plotted for each treatment group.

The listing of cGvHD will be provided. Onset of cGvHD (first event) will summarized. The proportion of subjects with different overall severity of cGvHD will be presented, as well as the proportion of patients with required systemic treatment for cGvHD.

Best Overall Response

Best overall response will be calculated based on the local investigators review of aGvHD assessment data. It will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

2.7.3 Handling of missing values/censoring/discontinuations

Refer to Section 2.5.3

2.8 Safety analyses

All safety analyses will be based on Safety Set, except that the summary of safety data during the crossover treatment phase will be based on Crossover Analysis Set. All listings and tables will be presented by treatment group. For safety evaluations (except for AE) during randomized treatment phase, the last available assessment on or before the date of start of randomized treatment is taken as the "baseline" assessment. For safety evaluations (except for AE) during crossover treatment phase, the last available assessment on or before the date of start of crossover treatment is taken as the "baseline" assessment.

Due to possible crossover from BAT to ruxolitinib arm after Day 28, imbalance in exposure between the two arms is expected. Therefore, safety summaries for the randomized treatment will be performed for the following periods, unless specified:

- Up to Day 31 (the upper bound of the Day 28 visit window);
- Up to the earlier of i) cutoff date, ii) end date of on-randomized-treatment period;

The on-randomized-treatment period is defined in Section 2.1.1.

For the summaries up to cutoff date, some may be presented by adjusting for the duration of randomized treatment period in patient-years where relevant.

2.8.1 Adverse events (AEs)

For reporting of AEs the overall observation period will be divided into mutually exclusive categories, including pre-treatment, on-treatment (randomized or crossover), post-treatment periods as defined in Section 2.1.1.

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

AEs will be summarized by number and percentage of patients having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding. A patient with multiple occurrences of an AE will be counted only once in the respective AE category. A patient with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

In AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on the descending frequency in the ruxolitinib arm.

The following adverse event summaries will be produced by treatment arm; overview of adverse events and deaths, AEs by SOC and PT, summarized by relationship (all AEs and AEs related to study treatment), seriousness (SAEs and non-SAEs), leading to treatment discontinuation, leading to dose interruption/adjustment, and leading to fatal outcome. In addition, a summary of serious adverse events with number of occurrences will be produced (an occurrence is defined as >1 day between start and prior end date of record of same preferred term).

2.8.1.1 AEs adjusted for patient duration of treatment period

In order to account for differences in exposure of the ruxolitinib arm relative to the BAT arm due to crossover from BAT to ruxolitinib after Day 28 visit, incidence rates of adverse events may be presented by adjusting for duration of treatment period in patient-years where relevant.

2.8.1.2 Adverse events of special interest / grouping of AEs

Data analysis of AESIs

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to compound ruxolitinib. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and/or PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely

fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad.

For each AESI (selection from the Case Retrieval Sheet), number and percentage of patients with at least one event of the AESI occurring during on-treatment period will be summarized. Summaries of AESIs will be provided by treatment arm, (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose adjustment/interruption, hospitalization, death etc.).

A listing of all grouping levels down to the MedDRA preferred terms used to define each AESI will be generated.

In addition to summarizing infections by CTCAE grade, they will also be summarized using infection severity (protocol Appendix 2) up to Day 31 (the upper bound of the Day 28 visit window), and cutoff date. Summary of infections by SOC, HLGT, HLT and PT up to Day 31, and cutoff date will be generated.

Proportion of patients developing second primary malignancies will be summarized for on-treatment period, post-treatment period and both.

2.8.2 Deaths

Separate summaries for on-treatment and all deaths (including post-treatment death) will be produced by treatment arm, system organ class and preferred term.

All deaths will be listed and post-treatment deaths will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened patients.

2.8.3 Laboratory data

On analyzing laboratory, data from all sources (central and local laboratories) will be combined. The summaries will include all assessments available for the lab parameter collected no later than end of the on-treatment periods (randomized or crossover).

Grading of laboratory values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher.

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

- Shift tables using CTC grades to compare baseline to the worst on-treatment value

The following listings will be produced for the laboratory data:

- Listings of all laboratory data, with CTC grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.
- Listing of all CTC grade 3 or 4 laboratory toxicities

In addition to the above mentioned tables and listings, other exploratory analyses, for example figures plotting time course of raw or change in laboratory tests over time or box plots may be produced.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

Not applicable.

2.8.4.2 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: height (cm), weight (kg), body mass index (kg/m2), body temperature (°C), pulse (beats per minute), systolic and diastolic blood pressure (mmHg).

Data handling

Vital signs collected on treatment will be summarized. Values measured outside of on treatment period will be flagged in the listings.

Data analysis

For analysis of vital signs the clinically notable vital sign criteria are provided in Table 2-5 below.

Table 2-5 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Weight (kg)	increase > 10% from Baseline	decrease > 10% from Baseline
Systolic blood pressure (mmHg)	>=180 with increase from baseline of >=20	<=90 with decrease from baseline of >=20
Diastolic blood pressure (mmHg)	>=105 with increase from baseline of >=15	<=50 with decrease from baseline of >=15
Pulse rate (bpm)	>=100 with increase from baseline of >25%	<=50 with decrease from baseline of > 25%
Body temperature	>= 39.1	-

The number and percentage of patients with notable vital sign values (high/low) will be presented by treatment arm.

A listing of all vital sign assessments will be produced by treatment arm and notable values will be flagged. In the listing, the assessments collected outside of on-treatment period will be flagged.

Additional Analyses

Time to first occurrence of grade 3 infection

Time to first occurrence of infection is defined as time from start of study treatment to the date of first occurrence of grade 3 infection severity per protocol Appendix 2, i.e. time in days is calculated as (start date of first occurrence of infection) – (start of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be **the earliest** of the following dates:

- death date
- end date of on-treatment period (randomized or crossover)
- data cut-off date
- withdrawal of informed consent date

Failure curves (ascending Kaplan-Meier curves) will be constructed by treatment arm. Median together with 95% confidence interval as well as 25th percentile and 75th percentile will be presented for each treatment arm.

A sensitivity analysis will be conducted considering deaths or end of treatment phase (randomized or crossover) without prior infection as competing risks. Cumulative incidence curve for time to grade 3 infection as well as estimates at 1, 2 and 6 months with 95% confidence intervals will be presented for each treatment group.

In addition, the median time to occurrence for the subset of patients who experienced infection will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.

2.9 Pharmacokinetic endpoints

The pharmacokinetic endpoints were reported in the primary analysis, and will not be reported in this second analysis. Further results will be presented in the full analysis.

Pharmacokinetic Analysis Set (PAS) will be used in all pharmacokinetic data analysis and PK summary statistics.

Plasma samples for extensive PK will be taken at Day 1 (start of treatment), at Day 7 (week 1) for the first 25 adult patients and all adolescents to characterize the PK after first dose, and at steady state by non-compartmental analysis. Additional PK samples will be taken at later visits for all patients to characterize exposure-efficacy, exposure-safety [redacted] as data allows. Concentrations will be expressed in mass per volume units.

PK parameters

The PK parameters that will be determined are shown in Table 2-6. The PK parameters of ruxolitinib will be calculated from the extensive PK data based on the non-compartmental methods using Phoenix WinNonlin® (Pharsight, Mountain View, CA) software. Additional PK parameters may be estimated as needed.

Table 2-6 Non-compartmental PK parameters for ruxolitinib

AUClast	The AUC from time zero to the last measurable concentration sampling time (tlast) (ng x hr/mL)
AUCinf	The AUC from time zero to infinity (ng x hr/mL)
AUCtau	The AUC calculated to the end of a dosing interval (12 hr) at steady-state (ng x hr/mL)
Cmax	The maximum (peak) observed plasma, blood, serum, or other body fluid drug concentration after single dose administration (ng/mL)
Ctrough	The observed plasma concentration obtained prior to administration of the next dose (pre-dose concentration) (mass x volume-1)
Tmax	The time to reach maximum (peak) plasma, blood, serum, or other body fluid drug concentration after single dose administration (thr)
Lambda_z	Smallest (slowest) disposition (hybrid) rate constant (hr^{-1}) may also be used for terminal elimination rate constant (hr^{-1})
T1/2	The elimination half-life associated with the terminal slope (λz) of a semi logarithmic concentration-time curve (hr).
CL/F	The total body clearance of drug from the plasma (L/hr)
Vz/F	The apparent volume of distribution during terminal phase (associated with λz) (L)
Racc	Accumulation ratio (AUC at steady state/AUC Day 1)

Descriptive statistics (n, arithmetic mean, CV% mean, standard deviation (SD), median, geometric mean, CV% geo-mean, minimum and maximum) will be presented for the ruxolitinib arm for Pharmacokinetic Analysis Set for all PK parameters defined in Table 2-6 except Tmax, where only n, median, minimum and maximum will be presented.

All individual PK parameters will be listed for the ruxolitinib arm using the Full Analysis Set.

The potential impact of severity of GI GvHD on ruxolitinib pharmacokinetic parameters will be explored by summarizing PK parameters by baseline lower GI stage and producing box plots of AUCinf, AUCtau and Cmax by lower GI stage.

PK parameters will also be summarized by patients who took concomitant CYP3A4 inhibitors versus those who didn't. Patients are defined as having taken concomitant CYP3A4 inhibitors if based on the concomitant medication summary they have taken CYP3A4 inhibitors on the day of or on the day before PK samples are taken. In addition to the summary tables box plots of Cmax, AUCinf and AUCtau will be produced by concomitant CYP3A4 inhibitor use. This will be repeated for Day 1 and Day 7 PK parameters. The list of CYP3A4 inhibitors will be documented in Data Handling Plan.

[redacted]

[redacted]

PK concentrations

Descriptive statistics (n, m (number of non-zero concentrations), arithmetic mean, CV% mean, SD, median, geometric mean, CV% geo-mean, minimum and maximum) for ruxolitinib concentration by lower GI stage and overall will be presented at each scheduled time point for the ruxolitinib arm for the Pharmacokinetic Analysis Set.

Individual concentration-time profiles for ruxolitinib concentrations with median will be displayed graphically for the ruxolitinib arm for Full Analysis Set on the semi-log view. In addition, the mean (+/- SD) and geometric mean concentration-time profiles for ruxolitinib by treatment over time will be displayed graphically for Pharmacokinetic Analysis Set on the linear and semi-log view.

All individual plasma ruxolitinib concentration data will be listed for the ruxolitinib arm for the Full Analysis Set.

Crossover concentrations will be included in all analyses on concentrations.

Handling of PK data below LLOQ or missing

All concentration values below the lower limit of quantitation (LLOQ) are set to zero by the Bioanalyst, and will be displayed in the listings as zero and flagged. LLOQ values will be treated as zero in any calculations of summary statistics, and treated as missing for the calculation of the geometric means and their CV%. The number of non-zero concentrations will also be reported in the summary statistics.

Missing values for any PK data will not be imputed and will be treated as missing.

2.10 PD and PK/PD analyses

Analysis of relationship between ruxolitinib exposure and efficacy/safety endpoints

This analysis will be described and reported separately to the CSR. The following are objectives for exposure-response analysis:

- Characterize the exposure-efficacy relationship of ruxolitinib in terms of exposure and efficacy response, with efficacy response defined as (overall response rate at Day 28 and durable response at Day 56, duration of response, overall survival at 6 months and any other relevant endpoints, and exposure defined plasmaconcentration, PK parameter or dose, as appropriate)

- Characterize the exposure-safety relationship of ruxolitinib in terms of exposure and safety response, where safety response is defined as various categorizations of AEs (frequency of;severity of AEs, AEs of interest to be defined prior to analysis or laboratory parameters).
- Average steady-state exposures and/or other PK parameters for the population will becomputed by the POP PK model accounting for dose modifications or dose interruptions up to the day prior to the day of assessments. Population PK derived parameters will be used for exposure-response analysis by appropriate methods

For duration of response and overall survival, if the primary endpoint is significant and sufficient events have accrued (i.e., at analysis time points after the primary analysis), a Cox regression model with appropriate patient demographic and prognostic factors as covariates and the log-trough level as a time dependent covariate will be fitted if appropriate. To account for dose adjustments between trough sampling time points the trough concentration will be adjusted by dividing by the actual dose prior to the trough in question and multiplying by the arithmetic mean dose since the last trough sample or start of dosing as applicable. The calculation of the arithmetic mean dose should count any dose interruptions as a zero dose for the days that no dose was given. Goodness of fit of the model will also be examined. The survival time should be calculated from the time of first dose rather than the time of randomization. The hazard ratio and 95% confidence interval for a two-fold increase in exposure will be displayed if appropriate. Kaplan-Meier curves may be used to summarize the data based on relevant quantiles of PK.

For incidence of specific AEs, Day 28 response (if the primary endpoint is significant) and Day 56 durable response (if the primary and key secondary endpoints are significant), logistic regression models may be used including log-average trough concentration and other demographic and prognostic covariates in the model as appropriate. The trough concentration will be similarly adjusted to account for dose adjustments as described above. The average trough concentration included in the model for Day 28 response will be the arithmetic mean of the Day 7, Day 14, Day 21 and Day 28 adjusted trough concentration. A similar average trough concentration will be calculated for the Day 56 response except the Day 42 and Day 56 adjusted concentration will be double weighted compared to the others because they are representative of concentration over 2 weeks rather than 1. For analysis of adverse events the last trough concentration prior to the AE will be included in the model. Other exposure measures may be considered if appropriate. The odds ratio and its 95% confidence interval for a two-fold increase in exposure will be displayed as appropriate.

2.11 Patient-reported outcomes

The FACT-BMT along with the EQ-5D-5L will be used to collect data on the patient's disease-related symptoms and health-related quality of life. Responses to the FACT-BMT and EQ-5D-5L will be generated in accordance with the respective scoring manual. Further details on the scoring of FACT-BMT and EQ-5D are given in Appendix 5.4.3.1 and Appendix 5.4.3.2.

These PRO instruments are planned to be administered on randomization day and every week during the first 2 months, and every 4 weeks thereafter until the end of treatment.

The baseline is defined as the last PRO assessment prior to or on the treatment start date .

FACT-BMT will not be used in patients under age 18 years.

Descriptive statistics (mean, standard deviation, median, Q1, Q3, minimum, and maximum) will be used to summarize the scored scales at each scheduled assessment time point for the FACT-BMT and EQ-5D using FAS. Additionally, change from baseline in the scores at the time of each assessment will be summarized. Patients with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

Missing items data in a scale will be handled based on each instrument manual. No imputation will be applied if the total or subscale scores are missing at a visit.

Subscale scores from the FACT-BMT and EQ-5D-5L will be displayed as mean profiles for each study arm, presented over time using time windows as described in Section 2.1.

[redacted]

[redacted]

2.14 Interim analysis

No formal interim analysis is planned for this trial.

Early PK analysis

To compare the exposure of 10 mg BID in the SR aGvHD population to the known exposure in MPN patients, the early extensive ruxolitinib PK data on the first 25 adult patients (including any adolescent patients randomized at that time) will be explored once available, also in the context of concomitant medications. However, there will be no comparison between the two treatment arms because PK data are not collected in BAT patients.

DMC safety data analysis

DMC will be instituted in this study and will review the safety data as outlined according the DMC charter. A separated set of SAP and TFL shells will be provided to detail the analysis.

3 Sample size calculation

The study with a total of 308 patients and 1:1 randomization (ruxolitinib vs. BAT) stratified on aGvHD grade (Grade II vs. Grade III vs. Grade IV) has 90% power to test for the primary endpoint (ORR at Day 28) and approximately 90% power to test for the key secondary endpoint (durable ORR at Day 56). The family wise α-level will be controlled at 0.025 overall for the two comparisons. Specifically, this study will claim to have achieved the efficacy objective when the primary endpoint ORR at Day 28 shows a significant treatment effect at one-sided $\alpha = 0.025$. Conditional on significance of the primary endpoint, the key secondary endpoint durable ORR at Day 56 will be tested at one-sided $\alpha = 0.025$.

3.1 Primary analysis

The sample size calculation is based on the primary variable ORR at Day 28. The hypotheses to be tested and details of the testing strategy are described in Section 2.5.2 and Section 3.

Based on Martin P. et al BBMT 2013, the ORR at Day 28 in the BAT arm is expected to be 58%. The stratum specific rates (Grade II 69%, Grade III 59%, Grade IV 50%) are obtained assuming that the ratio of aGvHD Grade II: III: IV is 0.2:0.4:0.4. It is expected that treatment with ruxolitinib will result in an 18% increase in the ORR, i.e., an expected odds ratio of 2.25 (which corresponds to an increase in ORR to 75%). Power for the CMH test, stratifying on aGvHD grade, was calculated using software package East V6. In order to ensure 90% power a total sample size of 308 patients are needed.

With a sample size of 154 patients in each treatment arm, an observed odds ratio greater than or equal to 1.63 would achieve statistical significance for the primary endpoint. Assuming that the observed response rates in Grades II/III/IV in BAT arm are 69%/59%/50% (overall 57%), observed response rates ≥ 78%/70%/62% (overall 68%) in the ruxolitinib arm would achieve statistical significance.

3.2 Power for analysis of key secondary variables

Durable ORR at Day 56, as the key secondary variable, will be formally statistically tested, provided that the primary endpoint ORR at Day 28 is statistically significant. The hypotheses

to be tested and details of the testing strategy are provided in Section 2.6.2. Based on [Van Groningen 2016], the durable ORR at Day 56 in the BAT arm is expected to be approximately 35%. The stratum specific rates (Grade II 45%, Grade III 36%, Grade IV 30%) are obtained assuming that the ratio of aGvHD Grade II: III: IV is 0.2:0.4:0.4. It is expected that treatment with ruxolitinib will result in a 20% increase in the ORR, i.e., an expected odds ratio of 2.25 (which corresponds to an increase in durable ORR to 55%). With these assumptions and sample size of 308 patients, the power for the key secondary endpoint is at least 90%.

With sample size 154 patients in each treatment arm, an observed odds ratio greater than or equal to 1.59 for durable ORR at Day 56 would achieve statistical significance. Assuming that the observed durable response rates in Grades II/III/IV in BAT arm are 45%/36%/30% (overall 35%), observed durable response rates ≥ 57%/47%/41% (overall 47%) in the ruxolitinib arm would achieve statistical significance.

4 Change to protocol specified analyses

Compared to the protocol version 00, 'discontinuation from randomized treatment' has been dropped as reason. Since some BATs have fixed duration which is less than 8 weeks, to avoid bias, discontinuation from randomized treatment will not be used to disqualify the response or durable response. Also duration of response will not be censored based on treatment discontinuation.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

The following rule should be used for the imputation of the dose end date for a given study treatment component:

Scenario 1: If the dose end date is completely missing and there is no EOT page and no death date, the patient is considered as on-going:

The patient should be treated as on-going and the cut-off date should be used as the dose end date.

Scenario 1 should not be applied for final CSR. All patients should have EOT page complete before the Database lock for Final CSR

Scenario 2: If the dose end date is completely or partially missing and the EOT page is available:

Please note that date of assessment on EOT eCRF might be very different from last date of dose.

Case 1: The dose end date is completely missing, and the EOT completion date is complete, then this latter date should be used.

Case 2: Only Year(yyyy) of the dose end date is available and yyyy < the year of EOT date:

Use Dec31yyyy

Case 3: Only Year(yyyy) of the dose end date is available and yyyy = the year of EOT date:

Use EOT date

Case 4: Both Year(yyyy) and Month (mm) are available for dose end date, and yyyy = the year of EOT date and mm < the month of EOT date:

Use last day of the Month (mm)

All other cases should be considered as a data issue and the statistician should contact the data manager of the study.

After imputation, compare the imputed date with start date of treatment, if the imputed date is < start date of treatment:

Use the treatment start date

Patients with missing start dates are to be considered missing for all study treatment component related calculations and no imputation will be made. If start date is missing then end-date should not be imputed.

5.1.2 AE, ConMeds and safety assessment date imputation

Table 5-1 Imputation of start dates (AE, CM) and assessments (LB, EG, VS)

Missing Element	Rule
day, month, and year	• No imputation will be done for completely missing dates

day, month	• If available year = year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY
	  ○ Else set start date = study treatment start date.
	• If available year > year of study treatment start date then 01JanYYYY
	• If available year < year of study treatment start date then 01JulYYYY

Missing Element	Rule
day	• If available month and year = month and year of study treatment start date then o If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY. o Else set start date = study treatment start date. • If available month and year > month and year of study treatment start date then 01MONYYYY • If available month and year < month year of study treatment start date then 15MONYYYY

Table 5-2 Imputation of end dates (AE, CM)

Missing Element	Rule (* if end date of the on-treatment period not > (death date, cut-off date, withdrawal of consent date))
day, month, and year	• Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	• If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period *
day	• If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

Any AEs and ConMeds with partial/missing dates will be displayed as such in the data listings.

Any AEs and ConMeds which are continuing as per data cut-off will be shown as 'ongoing' rather than the end date provided.

The above imputations are only used for analyses of time to and duration of AEs and concomitant medications.

5.1.2.1 Other imputations

Incomplete date of initial diagnosis of aGvHD

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

5.2 AEs coding/grading

Adverse events are coded using the Medical dictionary for regulatory activities (MedDRA) terminology.

The latest available MedDRA version at the time of the analyses should be used. The MedDRA version used should be specified in the footnote of relevant tables.

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 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).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTCAE grades are given in Novartis internal criteria for CTCAE grading of laboratory parameters. The latest available version of the document based on the underlying CTCAE version 4.03 at the time of analysis will be used. For laboratory tests where grades are not defined by CTCAE v4.03, results will be graded by the low/normal/high (or other project-specific ranges, if more suitable) classifications based on laboratory normal ranges.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. CTCAE Grade 5 is not defined for laboratory values. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

Imputation Rules

CTC grading for blood differentials is based on absolute values. However, this data may not be reported as absolute counts but rather as percentage of WBC.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

The following rules will be applied to derive the WBC differential counts when only percentages are available for a xxx differential

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

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4]

In order to apply the above formula, albumin values in g/L will be converted to g/dL by multiplying by 0.1), calcium values in mmol/L will be converted to mg/dL by dividing by 0.2495. For calculation of laboratory CTC grades 0 and 1, the normal range for derived corrected calcium is set to the same limits (in mg/dL) as for calcium.

CTC grades for the derived absolute WBC differential counts (neutrophils, lymphocytes) and corrected calcium will be assigned as described above for grading.

5.4 Statistical models

5.4.1 Primary analysis

The null hypothesis of equality of response rate in the two treatment arm will be tested against one-sided alternative. The statistical hypotheses are:

$$H_0: ORR_{rux} \leq ORR_{BAT} \quad vs. \quad H_1: ORR_{rux} > ORR_{BAT}$$

where ORR_{rux} is the probability of response in ruxolitinib and ORR_{BAT} is the probability of response in BAT.

The Cochran-Mantel-Haenszel chi-square test X^2_{CMH} (implemented again via SAS procedure FREQ with CMH option in the TABLES statement) will be used to test the difference in response rates between the treatment arms. The p-value corresponding to the CMH test for "general association" will be used which follows a Chi-square distribution with one degree of freedom.

If the sampling assumptions for chi-square test is not met, the exact Cochran-Mantel-Haenszel test will be used (implemented via SAS procedure MULTTEST). The test is performed by running a stratified version of the Cochran-Armitage permutation test [Armitage et al. 1969]. In studies with stratified randomization, the chi-square approximation is considered appropriate for the X^2_{CMH} statistics if the rule of Mantel and Fleiss [Mantel and Fleiss 1980] is satisfied.

Logistic Regression

Odds ratio will be used as a measure of association between treatment and response. The odds ratio will be derived from the logistic regression model (implemented using SAS procedure LOGISTIC, with treatment specified as an explanatory variable in the CLASS statement) which allows for including not only the stratification factor but also for adjustments for other covariates (both categorical and continuous). The odds ratio will be presented with 95% Wald confidence limits.

In cases where an exact test has been used to compare response rates, the odds ratio should be determined using exact logistic regression, and the odds ratio presented with exact 95% confidence limits. In these cases, SAS PROC LOGISTIC with EXACTONLY option will be used.

Confidence interval for response rate

Responses will be summarized in terms of percentage rates with $100(1-\alpha)\%$ confidence interval using exact binomial confidence interval (implemented using SAS procedure FREQ with EXACT statement for one-way table [Clopper and Pearson 1934].

Definition of new or additional systemic therapy for aGvHD

The data source to search for the new or additional systemic therapy would be from eCRFs of Prior and Concomitant Medication (CONMED) (with subcategory as treatment for aGvHD) and Dosage Administration Record (DAR).

Any of the following therapies represent new or additional systemic therapy for aGvHD:

1. Any new CNI therapy being initiated as 'treatment for aGvHD' after the baseline as recorded either on DAR or on CONMED, and never received prior to or at baseline
2. Any other systemic therapy (excluding CNI or systemic corticosteroid (methylprednisolone, prednisone and prednisolone)) being started as 'treatment for aGvHD' after the baseline and recorded on CONMED. Note: the therapy may have been taken prior to baseline but not at baseline.
3. Additional BAT at or after start of the initial BAT as recorded on DAR
4. Treatment with ruxolitinib after cross-over from BAT as recorded on DAR

5.4.2 Key secondary analysis

Same instructions as in Section 5.4.1.

5.4.3 Other secondary analysis

5.4.3.1 FACT-BMT

For FACT-BMT the subscale scores, the FACT-BMT total score and the FACT-BMT Trial Outcome Index (TOI) will be calculated.

The subscales are physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), functional well-being (FWB) and bone marrow transplant subscale (BMTS). For each subscale the corresponding score will be calculated based on the item response of the answered question according to the FACT-BMT Scoring Guide (Version 4). The scoring guide identifies those items that must be reversed before being added to obtain subscale totals. Negatively stated items are reversed by subtracting the response from "4". After reversing proper items, all subscale items are summed to a total, which is the subscale score.

If there are missing items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done by using the formula below:

Prorated subscale score

= [Sum of item scores] x [N of items in subscale]/[N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the relevant items were answered. Note that not all items (questions) are used to calculate the scores (see Scoring Guide, version 4). For all the scores the higher the score is the better the QOL is.

The FACT-BMT Trial Outcome Index (TOI, score range: 0-96) is calculated as

$$\underset{(\textbf{PWB score})}{\_\_\_\_\_} + \underset{(\textbf{FWB score})}{\_\_\_\_\_} + \underset{(\textbf{BMTS score})}{\_\_\_\_\_} = \_\_\_\_\_ = \textbf{FACT-BMT TOI}$$

The FACT-BMT total score (score range: 0-148) is then calculated as the sum of all un-weighted subscale scores:

(**PWB** score) + (**SWB** score) + (**EWB** score) + (**FWB** score) + (**BMTS** score) = **FACT-BMT Total score**

5.4.3.2 EQ-5D-5L

EQ-5D-5L consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Also the EQ-VAS is included. Each dimension has 5 levels (1-5).

The data were reported for all 5 dimensions. The proportion of reported problems for each level were reported. The changes from baseline were defined with a shift table. The below screenshot is from the Excel conversion document obtained from the EuroQol webpage on 5L Value Sets (http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html). This UK value will be used to convert the "health state" to an "index value" for all patients in all countries. For example, if the health state is 11125, then the index value would be .316 and so on.

EQ-5D-5L_Crosswalk_Value_Sets.xls [Read-Only] [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J	K	L
1		Health state	Denmark	France	Germany	Japan	Netherlands	Spain	Thailand	UK	US	Zimbabwe
2		5L profile	Denma	Franc	Germa	Jap	Netherlan	Spa	Thaila	U	U	Zimbab
3		11111	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.900
4		11112	0.856	0.929	0.999	0.829	0.845	0.932	0.814	0.879	0.876	0.864
5		11113	0.818	0.910	0.999	0.785	0.805	0.914	0.766	0.848	0.844	0.854
6		11114	0.671	0.769	0.809	0.761	0.592	0.731	0.660	0.635	0.700	0.792
7		11115	0.519	0.622	0.611	0.736	0.370	0.541	0.549	0.414	0.550	0.727
8		11121	0.859	0.910	0.910	0.814	0.874	0.910	0.780	0.837	0.861	0.846
9		11122	0.787	0.839	0.909	0.740	0.765	0.857	0.723	0.768	0.820	0.810
10		11123	0.768	0.820	0.909	0.721	0.736	0.843	0.708	0.750	0.809	0.800
11		11124	0.622	0.679	0.719	0.697	0.523	0.660	0.602	0.537	0.669	0.738
12		11125	0.469	0.532	0.521	0.672	0.301	0.470	0.491	0.316	0.524	0.673
13		11131	0.824	0.888	0.887	0.768	0.843	0.887	0.726	0.796	0.827	0.833
14		11132	0.770	0.817	0.887	0.718	0.745	0.838	0.701	0.740	0.806	0.797
15		11133	0.756	0.798	0.887	0.705	0.719	0.825	0.694	0.725	0.800	0.787
16		11134	0.609	0.657	0.697	0.681	0.506	0.642	0.588	0.512	0.661	0.725
17		11135	0.457	0.510	0.499	0.656	0.284	0.452	0.477	0.291	0.517	0.660

6 Reference

1. Anderson JR, Cain KC and Gelber RD (1983). Analysis of survival by tumor response. Journal of Clinical Oncology, 1(11), 710-719.
2. Armitage P, McPherson CK and Rowe BC (1969). Repeated significance tests on accumulating data. J .R. Statist. Soc. A, 132, 232-244.
3. Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41.
4. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrical, 26, 404-413.
5. FDA briefing Document Oncologic Drugs Advisory Committee Meeting, July 24, 2012, Evaluation of radiologic review of progression-free survival in Non-hematologic malignancies. http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/oncologicdrugsadvisorycommittee/ucm311141.pdf
6. Harris AC, Young R, Devine S, Hogan WJ, Ayuk F, Bunworasate U, Chanswangphuwana C, Efebera YA, Holler E, Litzow M, Ordemann R, Qayed M, Renteria AS, Reshef R, Wölfl M, Chen YB, Goldstein S, Jagasia M, Locatelli F, Mielke S, Porter D, Schechter T, Shekhovtsova Z, Ferrara JL, Levine JE. (2016) International, Multicenter Standardization of Acute Graft-versus-Host Disease Clinical Data Collection: A Report from the Mount Sinai Acute GVHD International Consortium. Biol Blood Marrow Transplant; 22(1): 4-10.
7. Korhonen P, Laird NM and Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC study (1999). Statistics in Medicine, 18, 2879-2897.
8. Mantel, N. and Joseph L Fleisś (1980). Minimum expected cell size requirements for the Mantel-Haenszel one degree-of-freedom Chi-Square Test and a related rapid procedure. American Journal of Epidemiology, 112, 129-134.
9. Robins JM and Tsiatis AA (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models (RPSFTM). Commun. Statist.-Theory Meth., 20(8), 2609-31.
10. Van Groningen, LF, Liefferink AM, de Haan AF, Schaap NP, Donnelly JP, Blijlevens NM, van der Velden WJ. (2016) Combination Therapy with Inolimomab and Etanercept for Severe Steroid-Refractory Acute Graft-versus-Host Disease Biol Blood Marrow Transplant. 22(1):179-82.

NOVARTIS

Clinical Development

INC424/ruxolitinib/JAKAVI

CINC424C2301 / NCT02913261

A phase III randomized open-label multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease after allogeneic stem cell transplantation

Statistical Analysis Plan (SAP)

Author:	Trial Statistician, [redacted]; Program Statistician, [redacted]

Document type:	Amended SAP Documentation
Document status:	Final
Release date:	31-Aug-2019
Number of pages:	55

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
12JUL2019	*8:00*	*Clarification of definition*	*Modified the definition of non-responder to be aligned within the compound*	*Sec 2.5, 2.6, 2.7*
12JUL2019	*8:00*	*Addition of definition*	*Added the definition of additional systemic therapy*	*Sec 5.4*
12JUL2019	*8:00*	*Clarification of definition*	*Updated the list of PDs for PPS patient classification*	*Sec 2.2*
26AUG019	*8:00*	*Clarification of definition*	*Added the scoring calculation rules for PRO*	*Sec 5.4*

Table of contents

List of abbreviations

AE	Adverse event
AESI	Adverse Event of Special Interest
aGvHD	Acute Graft vs. Host Disease
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
BAT	Best Available Therapy
bid	bis in diem/twice a day
BMI	Body Mass Index
BOR	Best Overall Response
BSA	Body Surface Area
cGvHD	chronic Graft vs. Host Disease
CIBMTR	Center for International Blood and Marrow Transplant Research
CMH	Cochran–Mantel–Haenszel
CMV	Cytomegalovirus
CR	Complete Response
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DAR	Dose Administration Record
DMC	Data Monitoring Committee
eCRF	Electronic Case Report Form
EFS	Event-Free Survival
FACT-BMT	Functional Assessment of Cancer Therapy - Bone Marrow Transplantation
FAS	Full Analysis Set
FFS	Failure-Free Survival
GvHD	Graft vs. Host Disease
HCT	Hematopoietic Cell Transplantation
IRT	Interactive Response Technology that includes Interactive Voice Response System and Interactive Web Response System
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NRM	Non Relapse Mortality
ORR	Overall Response Rate
OS	Overall Survival
PAS	Pharmacokinetic analysis set
PD	Pharmacodynamics
PK	Pharmacokinetics
PPS	Per-Protocol Set
PR	Partial response
PRO	Patient-reported Outcomes
QoL	Quality of Life
RAP	Report and Analysis Process

RPSFT	Rank-Preserving Structure Failure Time
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
WBC	White Blood Cells
SOC	System Organ Class
TFLs	Tables, Figures, Listings
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes all planned analyses for the Clinical Study Report (CSR) of study CINC424C2301, a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease (aGvHD Grade II-IV) after allogeneic stem cell transplantation.

The content of this SAP is based on protocol CINC424C2301 version 02. All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock of the study data.

1.1 Study design

This is a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory acute graft vs. host disease after allogeneic stem cell transplantation. Approximately 308 patients will be randomized to one of the following treatment arms in 1:1 ratio:

- Ruxolitinib
- BAT

Randomization will be stratified by the aGvHD grade (Grade II vs. Grade III vs. Grade IV).

Overall response rate (ORR) at Day 28, as assessed by local investigators' review of aGvHD response and using standard criteria [Harris 2016], is the primary endpoint in this study. Durable overall response rate at Day 56 is the key secondary endpoint.

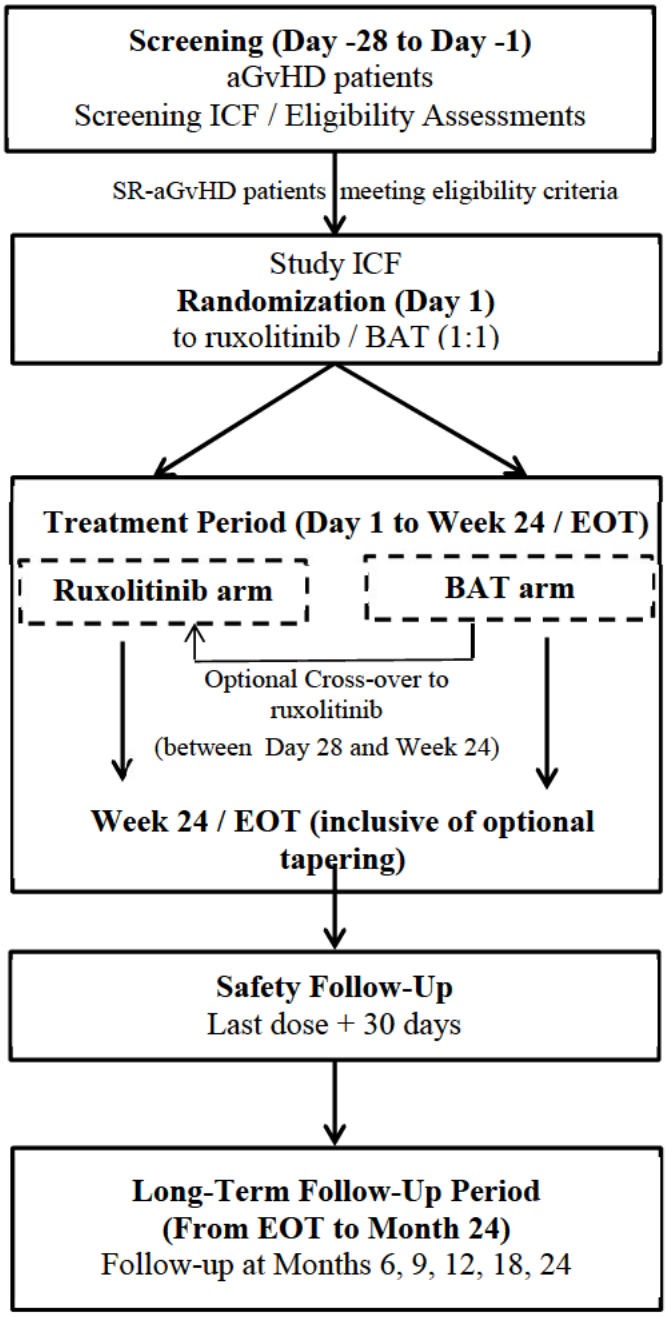
The primary analysis including the analysis on primary and key secondary endpoints will be performed after all patients have completed their Day 56 visit or have discontinued study. The primary analysis data will be summarized in the primary clinical study report (CSR).

Further analyses on secondary endpoints will be performed when all patients have completed 6 months treatment or discontinued from study participation earlier.

The final analysis will occur once all patients have completed the study (up to 24 months from randomization). All available data from all patients up to end of study, inclusive of OS, will be reported in a final CSR.

No formal interim efficacy analysis is planned in this study.

Figure 1-1 Schematic Study Design



1.2 Study objectives and endpoints

Table 1-1 Objectives and related endpoints

Objective	Endpoint	Analysis
Primary		**Refer to Section 2.5**
To compare the efficacy of ruxolitinib vs. Investigator's choice Best Available Therapy (BAT) in patients with Grade II-IV SR-aGvHD assessed by Overall Response Rate (ORR) at Day 28	Overall response rate (ORR) at Day 28 after randomization, defined as the proportion of patients in each arm demonstrating a complete response (CR) or partial response (PR) without requirement for additional systemic therapies for an earlier progression, mixed response or non-response. Scoring of response will be relative to the organ stage at the time of randomization.	CMH analyses, stratified by aGvHD grade
Key secondary endpoint		**Refer to Section 2.6**
To compare the rate of durable ORR at Day 56 between ruxolitinib and BAT	Proportion of all patients in each arm who achieve a complete response (CR) or partial response (PR) at Day 28 and maintain a CR or PR at Day 56.	CMH analyses, stratified by aGvHD grade
Other secondary endpoints		**Refer to Sections 2.7, 2.8, 2.9, 2.10, 2.11**
To estimate ORR at Day 14	Proportion of patients who achieved ORR (CR+PR) at Day 14.	
To assess Duration of response	Duration of response (DOR) is assessed for responders only and is defined as the time from first response until aGvHD progression or the date of additional systemic therapies for aGvHD. Onset of chronic GvHD, or death without prior observation of aGvHD progression are considered as competing risks.	

Objective	Endpoint	Analysis
To assess the cumulative steroid dose until Day 56	Weekly cumulative steroid dose for each patient up to Day 56 or end of treatment will be calculated.	
To assess Overall Survival (OS)	Overall survival, defined as the time from the date of randomization to the date of death due to any cause.	Probability of overall survival at Months 1, 2, 6, 12, 18 & 24 will be estimated from the Kaplan-Meier curves for each arm.
To assess Event-Free Survival (EFS)	Event-free survival, defined as the time from the date of randomization to the date of hematologic disease relapse/progression, graft failure, or death due to any cause.	Probability of EFS at Months 1, 2, 6, 12, 18 & 24 will be estimated from the Kaplan-Meier curves for each arm.
To assess Failure-Free Survival (FFS)	Failure-free survival, defined as the time from the date of randomization to date of hematologic disease relapse/progression, non-relapse mortality, or addition of new systemic aGvHD treatment.	Cumulative incidence of FFS at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering each event as a competing risk for the other two. Onset of chronic GvHD is considered as a competing risk.
To assess Non Relapse Mortality (NRM)	Non-relapse mortality (NRM), defined as the time from date of randomization to date of death not preceded by hematologic disease relapse/progression.	Cumulative incidence of NRM at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering hematologic disease relapse/progression as competing events.
To assess incidence of Malignancy Relapse/Progression (MR)	Malignancy Relapse/Progression (MR) (refer to protocol Appendix 4), defined as the time from date of randomization to hematologic malignancy relapse/progression. Calculated for patients with underlying hematologic malignant disease.	Cumulative incidence of MR at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering deaths not preceded by hematologic malignancy relapse/progression as competing events.
To measure the incidence of cGvHD	cGvHD, defined as the diagnosis of any cGvHD including mild, moderate, severe.	Cumulative incidence of cGvHD at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering hematologic disease relapse/progression and death without prior cGvHD as competing events.
To estimate the rate of Best Overall Response (BOR)	Proportion of patients who achieve OR (CR+PR) at any time point up to and including Day 28 and	

Objective	Endpoint	Analysis
	before the start of additional systemic therapy for aGvHD.	
To assess Pharmacokinetics (PK) of ruxolitinib in SR-aGvHD patients	Pharmacokinetic parameters of ruxolitinib after a single dose and at steady state. Cmax, AUClast, AUCinf, Ctrough, Racc and AUCtau. Other PK parameters are CL/F, Vz/F, Tmax and T1/2.	
To assess exposure-response relationship of ruxolitinib in SR-aGvHD	Pharmacokinetics (exposure) and efficacy (ORR, OS, or other relevant endpoints) relationship. Pharmacokinetics (exposure) and safety (AEs) relationship.	
To evaluate changes in Patient Reported Outcomes (PROs)	Change in FACT-BMT from baseline to each visit where measured. Change in EQ-5D-5L from baseline to each visit where measured.	
To evaluate the safety of ruxolitinib and Best Available Therapy	Safety and tolerability including myelosuppression, infections, and bleeding will be assessed by monitoring the frequency, duration and severity of Adverse Events including occurrence of any second primary malignancies, infections, by performing physical exams, and evaluating changes in vital signs from baseline, routine serum chemistry, hematology results and coagulation profile.	
[redacted]		[redacted]
[redacted]	[redacted]	

Objective	Endpoint	Analysis
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
	[redacted]	
	[redacted]	

2 Statistical methods

2.1 Data analysis general information

The final analysis will be performed by Novartis. SAS version 9.4 or later and/or R version 3.0.2 or later will be used to perform all data analyses and to generate tables, figures and listings.

Data included in the analysis

The analysis cut-off date for the primary analysis of study data will be established after all randomized patients have completed Day 56 visit or have discontinued treatment period or have discontinued study. All statistical analyses will be performed using all data collected in the database up to the data cutoff date. All data with an assessment date or event start date (e.g. vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.

All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'ongoing'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these events, the end date will not be imputed and therefore will not appear in the listings.

The analysis cutoff date for the final analysis of study data will be established when all patients have completed the study.

General analysis conventions

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

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

Quantitative data (e.g., age, body weight, etc.) will be summarized by appropriate

descriptive statistics (i.e. mean, standard deviation, median, minimum, and maximum) by treatment group.

2.1.1 General definitions

Investigational drug and study treatment

Investigational drug, will refer to the ruxolitinib only. Whereas, ***study treatment*** will refer to ruxolitinib and BAT.

Randomized treatment will refer to the study treatment received during the randomized treatment period. Up to Day 28 visit, more than one BAT regimen may be initiated as study

treatment. These regimens reported on the Dosage Administration Record (DAR) eCRF are considered randomized treatment. ***Crossover treatment*** will refer to the study treatment received during the crossover treatment period.

Date of first administration of randomized treatment

The date of first administration of randomized treatment is derived as the first date when a nonzero dose of randomized treatment was administered as per the DAR eCRF. The date of first administration of randomized treatment will also be referred as ***start of randomized treatment***.

Date of first administration of crossover treatment

The date of first administration of crossover treatment is derived as the first date when a nonzero dose of crossover treatment was administered as per the DAR eCRF. The date of first administration of crossover treatment (ruxolitinib) will also be referred as ***start of crossover treatment***.

Date of last administration of randomized treatment

The date of last administration of randomized treatment is defined as the last date when a nonzero dose of randomized treatment was administered as per DAR eCRF.

Date of last administration of crossover treatment

The date of last administration of crossover treatment is defined as the last date when a nonzero dose of crossover treatment (ruxolitinib) was administered as per DAR eCRF.

Study day

The study day, describes the day of the event or assessment date, relative to the reference start date.

The study day is defined as:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

The reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, PK, etc.), [redacted] and PRO is the start of randomized treatment.

The reference start date for all other, non-safety assessments (i.e., aGvHD assessment, survival, aGvHD progression, aGvHD response, underlying hematologic disease relapse/progression, etc.) is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Crossover study day

The crossover study day, describes the day of the event or assessment date, relative to the start of crossover treatment.

Crossover study day = date of event – start of crossover treatment + 1, if event is on or after the start of crossover treatment

Crossover study day = date of event – start of crossover treatment, if event precedes the start of crossover treatment

The crossover study day will be displayed in the data listings if an event starts on or after the start of crossover treatment.

Time unit

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

Baseline

Baseline for safety endpoints (except adverse events), [redacted] and patient reported outcomes is defined as the last assessment prior to or on the treatment start date.

For evaluations after cross over, the baseline is defined as the last assessment prior to or on the start date of crossover treatment.

Baseline for efficacy endpoints is defined as the last assessment or procedure conducted prior to or on the date of randomization date + 3 days, but no later than the treatment start date.

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

On-treatment assessment/event and observation periods

For adverse event reporting the overall observation period will be divided into four mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent at screening to the day before first administration of study treatment
2. ***on-randomized-treatment period***: from date of first administration of randomized treatment to 30 days after date of last actual administration of randomized treatment (including start and stop date) or end of randomized treatment per End of Randomized Treatment Disposition eCRF, whichever is later; for those patients who cross over from BAT to ruxolitinib, the period is from date of first administration of randomized treatment to earlier of (i) 30 days after date of last actual administration of randomized treatment or end of randomized treatment per End of Randomized Treatment Disposition eCRF, whichever is later, (ii) the day before the date of first administration of crossover treatment. Up to Day 28 visit, more than one BAT regimen may be initiated as study treatment. In this case, the last actual administration of randomized treatment refers to the last actual administration of the last BAT regimen reported on the DAR eCRF.

3. ***on-crossover-treatment period***: from date of first administration of crossover treatment to 30 days after date of last administration of crossover treatment (including start and stop date) or end of crossover treatment per End of Crossover Treatment Disposition eCRF, whichever is later.
4. ***post-treatment period***: starting at Day 31 after last administration of study treatment or the day after end of study treatment per end of treatment disposition eCRFs, whichever is later.

If dates are incomplete in a way that clear assignment to pre-, on-, post-treatment period cannot be made, then the respective data will be assigned to the on-randomized-treatment period.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). The data on-randomized-treatment period and on-crossover-treatment period will be summarized separately. In addition, a separate summary for death including on-treatment and post-treatment deaths will be provided. In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period (***treatment-emergent*** AEs).

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

Windows for multiple assessments

In order to summarize data collected over time (including unscheduled visits), the assessments will be time slotted. The following general rule will be applied in creating the assessment windows: If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If two assessments within a time window are equidistant from the target date, then the later of the two assessments will be used. If multiple assessments on the same date then the worst case will be used. Data from all assessments (scheduled and unscheduled), including multiple assessments, will be listed.

The following time windows are defined for descriptive summary on aGvHD assessment, [redacted] PROs and safety (Table 2-1) by visit. The end of treatment assessment will be mapped into the time points if collected within 7 days of the last dose intake.

Table 2-1 Time windows for aGvHD assessment, [redacted] PROs and safety assessment (lab, vital sign, etc.)

Time Window	Planned Visit Timing	Time Window Definition
On treatment		
Baseline[1]	Study Day 1[1]	≤ Study Days 1+3, no later than treatment start date
Baseline[2]	On or before Study Day 1[2]	≤ Study Day 1
Week 1	Study Day 7	Study Days 4 – 10
Week 2	Study Day 14	Study Days 11 – 17
Week 3	Study Day 21	Study Days 18 – 24
Week 4	Study Day 28	Study Days 25 – 31
Week 5	Study Day 35	Study Days 32 – 38

Week 6	Study Day 42	Study Days 39 – 45
Week 7	Study Day 49	Study Days 46 – 52
Week 8	Study Day 56	Study Days 53 – 59
Week 12	Study Day 84	Study Days 71 – 98
Week 16	Study Day 112	Study Days 99 – 126
Week 20	Study Day 140	Study Days 127 – 154
Week 24	Study Day 168	Study Days 155 – 182
Safety follow-up	30 days after last dose	Last dose date + 30

Baseline[1] for aGvHD assessment (efficacy);
Baseline[2] for safety assessments [redacted] (safety);
Study Day 1[1] = randomization date;
Study Day 1[2] = start date of randomized treatment;
Crossover Study Day 1 = start date of crossover treatment
EOT (randomized treatment or crossover treatment) assessments are mapped to the time points.
Safety follow-up is a separate time point for PROs and safety assessments.

Last contact date

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

Table 2-2 Last contact date data sources

Source data	Conditions
Date of Randomization	No Condition
Last date patient was known to be alive from Survival Follow-up page	Patient status is reported to be alive, lost to follow-up or unknown.
Start/End dates from further aGvHD therapy	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
End of treatment date from end of treatment page	No condition.
- aGvHD assessment date - any specific efficacy assessment date if available (e.g., cGvHD assessment, graft failure assessment, hematologic disease relapse/progression assessment)	Evaluation is marked as 'done'.
Laboratory/PK collection dates/[redacted]	Sample collection marked as 'done'.
Vital signs date	At least one non-missing parameter value
Start/End dates of AE	Non-missing verbatim term

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date . The cut-off date will not be used for last contact date, unless the patient was seen or contacted on that date. No date post cut-off date will be used. Completely 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. Partial date imputation is allowed for event (death)/censoring coming from 'Survival information' eCRF.

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

2.2 Analysis sets

Full Analysis Set

The Full Analysis Set (FAS) comprises all patients to whom study treatment has been assigned by randomization. According to the intent to treat principle, patients will be analyzed according to the treatment and strata, they have been assigned to during the randomization procedure.

Per-Protocol Set

The Per-Protocol Set (PPS) consists of a subset of the patients in the FAS who are compliant with requirements of the clinical study protocol.

The following list of protocol deviations will lead to exclusion of the patient from the Per-Protocol Set:

- Not steroid refractory aGvHD (PDID: INCL09 per SSD v22.0)
- More than one prior systemic therapy for the treatment of aGvHD other than corticosteroids +/- CNI (prophylaxis or treatment) (PDID: EXCL01 per SSD v22.0)
- Missing or incorrect aGvHD grade at randomization (PDID: OTH02, INCL07 per SSD v22.0)
- Taking any prohibited medication as specified in this protocol after start of study treatment and before end of study treatment (PDID: COMD01, COMD02 per SSD v22.0)
- Study treatment received different from treatment arm assigned by randomization (PDID: TRT02 per SSD v22.0)

Safety Set

The Safety Set includes all patients who received at least one dose of study treatment. Patients will be analyzed according to the study treatment received, where treatment received is defined as the randomized treatment if the patient took at least one dose of that treatment or the first treatment received if the randomized treatment was never received during the randomized treatment period.

Crossover Analysis Set

The Crossover Analysis Set (CAS) comprises all patients randomized to and received BAT who then crossed over and received at least one dose of ruxolitinib. This analysis set will be used for all analyses for crossover patients.

Pharmacokinetic Analysis Set

The Pharmacokinetic Analysis Set (PAS) includes all patients who provide at least one evaluable PK concentration. For a concentration to be evaluable, patients are required to:

- Take a dose of ruxolitinib prior to sampling.
- For pre-dose samples, do not vomit within 2 hours after the dosing of ruxolitinib prior to sampling; for post-dose samples, do not vomit within 2 hours after the dosing of ruxolitinib.

The PAS will be used for NCA analysis for patients where extensive PK sampling is obtained [redacted]

Patient Classification:

Patients may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific patient classification rules defined in Table 2-3.

Table 2-3 Patient classification based on protocol deviations and non-PD criteria

Analysis set	Protocol deviations leading to exclusion	Non protocol deviation leading to exclusion
FAS	INCL03	Not applicable
Safety Set	INCL03	No dose of study treatment
Per-Protocol Set	INCL03, INCL07, INCL09, EXCL01, TRT02, , OTH02, During randomized treatment period: COMD01, COMD02	No dose of study treatment
Crossover Analysis Set	INCL03	No dose of ruxolitinib
PK Analysis Set	INCL03	No dose of ruxolitinib, No evaluable PK concentration

Note: Based on CINC424C2301_Study Specification Document version 22.0.
INCL03 - Written Study informed consent /assent not obtained.

Withdrawal of Informed Consent

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

2.2.1 Subgroup of interest

Efficacy

The primary efficacy endpoint will be summarized by the following subgroups to examine the homogeneity of treatment effect provided that the primary efficacy analysis based on the FAS is statistically significant :

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race
- Region (Europe (including Australia and Canada), Asia excluding Japan, Japan)
- Acute GvHD grade (Grade II, III, IV)
- Source of grafts (related, not related)
- Criteria for SR-aGvHD (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper)
- Prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/- other systemic aGvHD treatment,)

No formal statistical test of hypotheses will be performed for the subgroups, only point estimate of the treatment effect and 95% confidence intervals will be provided (see Section 2.5 for further analysis details). The objective of the efficacy subgroup analysis is to demonstrate homogeneity of treatment effect in the above subgroups.

Safety

Key safety analyses will be repeated on the Safety Set in the following subgroups:

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race

The objective for carrying out these subgroup analyses is to identify potential safety issues that may be limited to or more commonly observed in a subgroup of patients. The following summaries will be presented by subgroup:

- AEs, irrespective of causality, by primary system organ class and preferred term
- AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- Serious AEs, irrespective of causality, by primary system organ class and preferred term
- Serious AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- On-treatment deaths, by primary system organ class and preferred term

Adolescent patients

In order to be able to make a separate risk/benefit assessment for the adolescent patients, besides the above planned subgroup analyses for this age group (12-<18 years), data of demographics and exposure will be presented.

Japanese patients

Subgroup analyses will also be performed for the patients treated in Japan. No selection will be done on the basis of ethnicity, the purpose being to evaluate the population of patients living in

Japan, not a specific ethnic set of patients. The analysis will be done the same way as in region subgroup specified above.

2.3 Patient disposition, demographics and other baseline characteristics

The Full Analysis Set (FAS) will be used for all baseline and demographic summaries and listings unless otherwise specified. Summaries will be reported by treatment arm and for all patients and listings will be reported by treatment arm to assess baseline comparability. No inferential statistics will be provided.

Basic demographic and background data

All demographic and baseline disease characteristics data will be summarized and listed by treatment arm. Categorical data (e.g. gender, age groups: 12- <18 vs. 18-65 vs. >65 years, race, ethnicity) will be summarized by frequency counts and percentages; the number and percentage of patients with missing data will be provided. Continuous data (e.g. age, weight, height, body mass index) will be summarized by descriptive statistics (N, mean, median, standard deviation, Q1, Q3, minimum and maximum).

Baseline stratification factors

The number (%) of patients in each stratum (Grades II, III, IV) based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS. Discordances between the stratum recorded in IRT at the time of randomization and the actual stratum recorded in the clinical database through the data collected on eCRF will be cross-tabulated and listed.

Diagnosis and extent of disease

Summary statistics will be tabulated for diagnosis and extend of disease in underlying disease, stem cell transplant and acute GvHD.

For underlying disease, the analysis will include the following: primary diagnosis category and subcategory, details of primary diagnosis, time since diagnosis of underlying disease, CIBMTR risk assessment.

For transplant related disease history, the analysis will include the following: conditioning regimen type, total HCT-specific comorbidity index score, time since transplant, time from diagnosis of underlying disease to transplant, stem cell type, cytomegalovirus status, donor information including age, gender, HLA typing method, HLA match score, source of grafts (related/unrelated), CMV status, T-cell depleted (Y/N), total nucleated cell dose.

For aGvHD disease history, the analysis will include the following: time since diagnosis of aGvHD grade ≥2, aGvHD grade when diagnosis of grade ≥2, steroid refractory aGvHD criteria met (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper), prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/-other systemic aGvHD treatment,), time from diagnosis of aGvHD grade ≥2 to steroid refractory, time since steroid refractory aGvHD, aGvHD grade at randomization, aGvHD organ involvement, steroid dose at randomization.

Medical history

Medical history and ongoing conditions, including underlying disease conditions and symptoms entered on eCRF will be summarized and listed by treatment arm. The summaries will be presented by primary system organ class (SOC), preferred term (PT) and treatment arm. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

Other

Intended BAT strategy in case of SR aGvHD prior to randomization will be summarized by treatment arm.

All data collected at baseline including child bearing potential will be listed.

2.3.1 Patient disposition

Enrollment by country and center will be summarized for all screened patients and also by treatment arm. The number (%) of randomized patients will be presented overall and by treatment group. The number (%) of screened and not-randomized patients and the reasons for screening failure will also be displayed. The number (%) of patients in the FAS who are still on treatment, who discontinued the treatment phases (randomized treatment, crossover treatment) as well as the reason for discontinuation, and the survival follow-up will be presented overall and by treatment group.

Protocol deviations

The number (%) of patients in the FAS with any protocol deviation will be tabulated by deviation category (as specified in the study Data Handling Plan) overall and by treatment group for the FAS. All protocol deviations will be listed.

Analysis sets

The number (%) of patients in each analysis set (defined in Section 2.2) will be summarized by treatment group and stratum.

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

2.4.1 Study treatment / compliance

Duration of exposure in days to ruxolitinib and BAT will be summarized by means of descriptive statistics. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. Duration of exposure to each BAT regimen (anti-thymocyte globulin (ATG), extracorporeal photopheresis (ECP), mesenchymal stromal cells (MSC), low-dose methotrexate (MTX), mycophenolate mofetil (MMF), mTOR inhibitors (everolimus or sirolimus), etanercept, or infliximab) will be summarized using the same approach.

Actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized for the ruxolitinib arm. Patients randomized to Investigator's choice of BAT will receive various different categories of therapy.

The number (%) of patients who have dose changes or interruptions, and the reasons, will be summarized for ruxolitinib group.

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

The Safety Set and Crossover Analysis Set will be used for all summaries on randomized treatment and crossover treatment, respectively. The Safety Set will be used for listings of study treatment.

Duration of exposure to study treatment

Duration of exposure to randomized treatment (days) = (last date of exposure to randomized treatment) – (date of first administration of randomized treatment) + 1.

Duration of exposure to crossover treatment (days) = (last date of exposure to crossover treatment) – (date of first administration of crossover treatment) + 1.

The last date of exposure to study treatment is the latest of the last dates of exposure to study treatment (see Table 2-4).

Table 2-4 Definition of last date of exposure of study treatment

Scenario	Definition of last date of exposure of study treatment	Example
Scenario 1: Study treatment with a periodical administration	The planned end date of the last period in which the last non-zero dose of the study treatment was last administered. Note : If the patient changed to another BAT, died or was lost to follow-up before the derived last date, the last date of exposure to study treatment is the date prior to the start of the next BAT, date of death or the date of last contact, respectively. If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	Example 1: In a once-a-week administration, the last date of exposure is the date of last administration + 6 days. Example 2: In a twice-a-week administration, the last date of exposure is the date of last administration + 3 days.
Scenario 2: Study treatment with daily/IV administration	Date of last administration of a non - zero dose of the study treatment.	Example 3: A patient had a permanent discontinuation of the study treatment 06Jan2017 after being put on a temporary interruption since 01Jan2017. In

		this case the last date of exposure is- 31Dec2016.
Scenario 3: Study treatment as an antibody	Date of last administration of a non -zero dose of the study drug + number of days antibody persists in vivo – 1 day. Note : If the patient changed to another BAT, died or was lost to follow-up before the derived last date, the last date of exposure to study treatment is the date prior to the start of the next BAT, date of death or the date of last contact, respectively. If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	Example 4: For a study treatment which antibody persists in vivo for 28 days, the last date of exposure is the date of last administration + 28 days – 1 day.

Summary of duration of exposure of study treatment in days will include categorical summaries (based on clinically meaningful time intervals) and continuous summaries (i.e. mean, standard deviation etc.) using appropriate units of time.

Duration of exposure in patient-years

The duration of exposure in patient-years is a total of the duration of exposure in years from all the patients in a treatment group. It will be calculated for randomized treatment (by treatment group) and crossover treatment, respectively.

Duration of treatment period

Duration of randomized treatment period (days) = end date of on-randomized-treatment period – date of first administration of randomized treatment + 1

Duration of crossover treatment period (days) = end date of on-crossover-treatment period – date of first administration of crossover treatment + 1

The on-randomized-treatment period and on-crossover-treatment period are defined in Section 2.1.1.

The duration of randomized treatment period in days for ruxolitinib and BAT will be summarized by means of descriptive statistics. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The duration of crossover treatment period will be summarized similarly.

Duration of treatment period in patient-years

The duration of treatment period (randomized and crossover) in patient-years is a total of the duration of treatment period in years from all the patients in a treatment group. It will be calculated for randomized treatment (ruxolitinib vs. BAT) and crossover treatment separately.

Cumulative dose

The **planned cumulative dose** for ruxolitinib refers to the total planned dose as per the protocol (10 mg bid) up to the last dose date.

The **actual cumulative dose of randomized ruxolitinib** refers to the total actual dose of randomized ruxolitinib as documented in the DAR eCRF.

The **actual cumulative dose of crossover ruxolitinib** refers to the total actual dose administered, over the duration for which the patient is on the crossover ruxolitinib as documented in the DAR eCRF.

For patients who did not take any drug the cumulative dose is by definition equal to zero.

Dose intensity and relative dose intensity of ruxolitinib

Dose intensity (DI) **of randomized ruxolitinib** for patients with non-zero duration of exposure is defined as follows:

DI (mg / day) = Actual cumulative dose (mg) of randomized ruxolitinib / Duration of exposure to randomized ruxolitinib (days).

Dose intensity (DI) **of crossover ruxolitinib** for patients with non-zero duration of exposure is defined as follows:

DI (mg / day) = Actual cumulative dose (mg) of crossover ruxolitinib / Duration of exposure to crossover ruxolitinib (days).

For patients who did not take any drug the DI is by definition equal to zero.

Planned dose intensity (PDI) is defined as follows:

PDI (mg / day) = Planned Cumulative dose (mg) / Duration of exposure (days).

The protocol planned starting dose for ruxolitinib is 10 mg BID.

Relative dose intensity (RDI) is defined as follows:

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

DI and RDI will be summarized for randomized and crossover ruxolitinib treatment, separately.

The actual cumulative dose, DI and RDI up to Day 28 visit, Day 56 visit and last date of exposure to study treatment (randomized or crossover) will be summarized.

The number (%) of patients at total daily dose 5 mg, 10 mg, 15 mg and 20 mg will be summarized at Day 28 visit, Day 56 visit and last date of exposure to study treatment.

Dose reductions, interruptions or permanent discontinuations

The number of patients who have dose changes, or interruptions, and the reasons, will be summarized for ruxolitinib (randomized and crossover treatment separately). The number of patients who have dose permanent discontinuations and the reasons, will be summarized by treatment group.

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

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

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

2.4.2 Prior, concomitant and post therapies

2.4.2.1 Prior aGvHD treatment

The number and percentage of patients who received any prior aGvHD treatment (medications and procedures) will be summarized by lowest ATC class, preferred term and treatment arm.

Listings will be produced for prior aGvHD treatment.

The above analyses will be performed using the FAS.

2.4.2.2 Prior prophylaxis

The number and percentage of patients who received any prophylaxis prior to randomization will be summarized by lowest ATC class, preferred term and treatment arm using FAS.

Listings will be generated for prophylaxis.

2.4.2.3 Systemic corticosteroid

The duration of exposure will be summarized for on-randomized-treatment period and on-crossover-treatment period separately. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The actual cumulative dose, dose intensity and relative dose intensity (relative to the starting dose of corticosteroids) will be summarized up to Day 28 visit, Day 56 visit and end of on-treatment period. These analyses will be based on Safety Set. The data will be reported on the DAR eCRF through the end of randomized treatment per EOT Disposition eCRF, and on the Prior and Concomitant Medications eCRF after the end of randomized treatment. The list of systemic corticosteroid will be documented in Data Handling Plan.

2.4.2.4 Calcineurin inhibitors (CNIs) during study treatment

The duration of exposure will be summarized for CNIs (cyclosporine or tacrolimus) during on-randomized-treatment period and on-crossover-treatment period, respectively. It will also be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. These analyses will be based on Safety Set. The data will be reported on the DAR eCRF through the end of randomized treatment per EOT Disposition eCRF, and on the Prior and Concomitant Medications eCRF after the end of randomized treatment. The list of CNIs will be documented in Data Handling Plan.

2.4.2.5 Additional systemic aGvHD therapy

New additional systemic aGvHD therapy (medications and procedures) since start of study treatment will be listed and summarized by lowest ATC class, preferred term, overall and by treatment group by means of frequency counts and percentages using FAS.

2.4.2.6 Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) other than the study treatment administered to a patient coinciding with the study treatment period. Concomitant therapy include medications (other than study drugs) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system and summarized by lowest ATC class and preferred term using frequency counts and percentages. Surgical and medical procedures will be coded using MedDRA and summarized by SOC and preferred term. The summaries for randomized treatment phase using Safety Set will include:

- Medications starting on or after the start of randomized treatment but no later than end of on-randomized-treatment period; and
- Medications starting prior to start of randomized treatment and continuing after the start of randomized treatment.

The summaries for crossover treatment phase using Crossover Analysis Set will include:

- Medications starting on or after the start of crossover treatment but no later than end of on-crossover-treatment period; and
- Medications starting prior to start of crossover treatment and continuing after the start of crossover treatment.

All concomitant therapies will be listed using Safety Set. Any concomitant therapies starting and ending prior to the start of randomized treatment or starting beyond end of on-randomized-treatment period if not crossed over, or starting beyond end of on-crossover-treatment period if crossed over, will be flagged in the listing.

The prohibited concomitant medications will be summarized by lowest ATC class and preferred term up to the end of on-randomized-treatment and on-crossover-treatment periods, respectively. The list of prohibited medications will be provided and updated regularly by clinical team according to the clinical database review.

In addition, a subset of concomitant medications i.e. transfusions (red blood cells and platelets) will be grouped and summarized by treatment group.

2.5 Analysis of the primary objective

The primary objective of the study is to compare the overall response rate (ORR) at Day 28 between the ruxolitinib arm and BAT arm in steroid refractory aGvHD patients.

2.5.1 Primary endpoint

ORR at Day 28 is defined as the proportion of patients with complete response (CR) or partial response (PR) at Day 28 according to standard criteria [Harris 2016]. Note that response is relative to the assessment of aGvHD at randomization.

- **Complete response** is defined as a score of 0 for the aGvHD grading in all evaluable organs that indicates complete resolution of all signs and symptoms of aGvHD in all evaluable organs without administration of additional systemic therapies for any earlier progression, mixed response or non-response of aGvHD.
- **Partial response** is defined as improvement of 1 stage in 1 or more organs involved with aGvHD signs or symptoms without progression in other organs or sites without administration of additional systemic therapies for an earlier progression, mixed response or non-response of aGvHD.
- **Lack of response** is defined as no response, mixed response, or progression.
  - **No response** is defined as absence of improvement in any organ involved by aGvHD, without worsening in any involved organ.
  - **Mixed response** is defined as improvement of at least 1 stage in the severity of aGvHD in at least one organ accompanied by progression in another organ or development of signs or symptoms of aGvHD in a new organ.
  - **Progression** is defined as worsening in 1 or more organs by 1 or more stages without improvement in any involved organ

A flare in aGvHD is defined as any increase in signs or symptoms of aGvHD after an initial response (CR or PR). However, a flare may not lead to progression or additional systemic therapy. Only flares in GvHD that require new additional systemic therapy, will be considered aGvHD flare failure. Patients who fail corticosteroid taper fulfilling either one of the following criteria should initiate additional systemic therapy:

- Requirement for an increase in the corticosteroid dose to methylprednisolone ≥2 mg/kg/day (or equivalent prednisone dose ≥2.5 mg/kg/day), OR
- Failure to taper the methylprednisolone dose to <0.5 mg/kg/day (or equivalent prednisone dose <0.6 mg/kg/day) for a minimum 7 days.

A patient will not be considered a responder at Day 28 if any of the following events occurs:

- Missing aGvHD assessment at baseline or Day 28
- No CR or PR at Day 28
- Additional systemic therapy for aGvHD prior to Day 28

Compared to the protocol version 00, 'discontinuation from randomized treatment' has been dropped as reason. Since some BATs have fixed duration which is less than 8 weeks, to avoid bias, discontinuation from randomized treatment will not be used to disqualify the response or durable response. Also duration of response will not be censored based on treatment discontinuation.

Acute GvHD will be assessed according to standard criteria [Harris 2016], as described in protocol Appendix 1. Grade will be calculated based on the staging of the organs and recorded on the eCRF by the Investigator. The Investigator reported grade will be used for randomization

and reported grade and response for all analyses. Grade and response will be calculated by the sponsor for the purposes of data review only and sensitivity analysis.

2.5.2 Statistical hypothesis, model, and method of analysis

The primary efficacy analysis is the comparison of ORR at Day 28 between the two treatment arms. The following statistical hypotheses will be tested to address the primary efficacy objective:

H_0: $ORR_{rux} \leq ORR_{BAT}$ vs. H_1: $ORR_{rux} > ORR_{BAT}$

where ORR_{rux} and ORR_{BAT} are the overall response rates at Day 28 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., aGvHD Grade II vs III vs. IV), will be used to compare ORR between the two treatment groups, at the one-sided 2.5% level of significance.

The primary efficacy variable, ORR at Day 28, will be analyzed at the time when all patients have completed their Day 56 visit or discontinued earlier. The primary analysis will be performed on FAS according to ITT principle. ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. One-sided p-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

2.5.3 Handling of missing values/censoring/discontinuations

Patients with missing assessments that prevent the evaluation of the primary endpoint will be considered non-responders on that treatment arm. This includes missing aGvHD response assessments at baseline and Days 28, 56.

The following analysis windows (also in Table 2-1) will be applied to the target day for assessments on overall response, where target day for Week X is X*7.

Baseline assessment is the last aGvHD assessment prior to or on the date of randomization (Day 1). A tolerance of up to 3 days from randomization will be considered in the analysis, but the baseline assessment will not be later than the date of treatment start.

Weeks 1, 2, 3, 4, 5, 6, 7, 8: -3 days/+3 days

Weeks 12 to 24: -13 days/+14 days

The analysis windows for assessments after crossover is similar, except that the baseline is the last aGvHD assessment prior to or on Crossover Day 1 (date of first administration of crossover treatment).

No data imputation will be applied.

2.5.4 Supportive analyses for ORR at Day 28

To evaluate the treatment effect on each organ, shift tables of aGvHD stage by organ and treatment group will be produced to compare baseline to the Day 28 value.

As supportive analyses, the primary endpoint will also be evaluated based on the primary analysis source (i.e., investigator assessment) with the same analysis conventions as for the primary efficacy analysis using all patients in PPS.

If the primary analysis is statistically significant, subgroup analyses to assess the homogeneity of the treatment effect across demographic and baseline disease characteristics will be performed on the FAS. The subgroups include but may not be limited to:

- Age group (12- <18, 18-65, >65 years)
- Gender
- Race
- Region (Europe (including Australia and Canada), Asia excluding Japanok, Japan)
- Acute GvHD grade (Grade II, III, IV)
- Source of grafts (related, unrelated))
- Criteria for SR-aGvHD (progression after at least 3 days, failure to achieve a response after 7 days, flare failure during taper)
- Prior aGvHD therapy (steroid +/- CNI, steroid +/- other systemic aGvHD treatment, steroid +/- CNI +/- other systemic aGvHD treatment-)

For each of the subgroups, the following analyses will be performed:

- Proportion of patients with ORR using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]
- Odds ratio with 95% CI using a logistic regression model with treatment and stratification factors as covariate

[redacted] Forest plot (n, odds ratio, 95% CI) will be produced to graphically depict the treatment effect estimates in different subgroups. No inferential statistics (p-values) will be produced for the subgroups.

A supportive analysis using FAS will be conducted using logistic regression model to estimate the treatment effect adjusting for key baseline and prognostic factors. The model may include the followings covariates: age, gender, race, aGvHD grade, source of grafts, criteria for SR-aGvHD, prior aGvHD therapy in addition to treatment as one of the covariates. Goodness of fit of the model will be examined.

As a sensitivity analysis using FAS to assess the impact of stratification, the two treatment groups will be compared using Fisher's exact test. ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934].

2.5.5 ORR at Crossover Day 28

ORR at Crossover Day 28 is defined as the proportion of crossover patients with complete response (CR) or partial response (PR) at Crossover Day 28 according to standard criteria [Harris 2016]. ORR will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] based on the CAS using local investigators'

review of aGvHD assessment data. Note that response is relative to the last assessment of aGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a responder at Crossover Day 28 if any of the following events occurs:

- Missing aGvHD assessment at Crossover baseline or Crossover Day 28
- No CR or PR at Crossover Day 28
- Additional systemic therapy for aGvHD prior to Crossover Day 28

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Crossover baseline to Crossover Day 28.

2.6 Analysis of the key secondary objective

The key secondary objective of the study is to determine whether treatment with ruxolitinib has better durable ORR at Day 56 compared with BAT.

2.6.1 Key secondary endpoint

Durable ORR at Day 56 is defined as the proportion of all patients in each arm who achieve a complete response (CR) or partial response (PR) at Day 28 and maintain a CR or PR at Day 56. If a patient is a CR at Day 28 and a PR at Day 56, he/she will be considered as a durable responder. A patient will not be considered a durable responder at Day 56 if any of the following events occurs:

- Not a responder at Day 28
- Missing aGvHD assessment at Day 56
- No CR or PR at Day 56.
- Additional systemic therapy for aGvHD prior to Day 56

Durable ORR will be calculated based on the FAS using local investigators review of aGvHD assessment data.

The patients randomized to BAT who meet cross-over criteria and cross-over to ruxolitinib are considered to have the additional systemic therarpy for aGvHD, and will not be considered as a responder afterwards.

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Day 28 to Day 56 value for the patients who achieved PRs at Day 28.

2.6.2 Statistical hypothesis, model, and method of analysis

The following statistical hypotheses will be tested to address the key secondary efficacy objective:

H_0: $DORR_{rux} \leq DORR_{BAT}$ vs. H_1: $DORR_{rux} > DORR_{BAT}$

where $DORR_{rux}$ and $DORR_{BAT}$ are the durable overall response rates at Day 56 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., aGvHD Grade II vs III vs. IV), will be used to compare

the durable ORR between the two treatment groups, at the one-sided 2.5% level of significance, if the primary endpoint is significant.

The key secondary efficacy variable, durable ORR at Day 56, will be analyzed at the time when all patients have completed their Day 56 visit or discontinued earlier. The analysis will be based on FAS according to ITT principle. Durable ORR will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. P-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

The durable ORR at Day 56 will be tested hierarchically. That is, if the ORR at Day 28 is statistically significant, the durable ORR at Day 56 will be tested. If the ORR at Day 28 is not statistically significant, the durable ORR at Day 56 will not be tested.

2.6.3 Handling of missing values/censoring/discontinuations

Refer to Section 2.5.3

2.6.4 Durable ORR at Crossover Day 56

Durable ORR at Crossover Day 56 is defined as the proportion of all crossover patients who achieve a complete response (CR) or partial response (PR) at Crossover Day 28 and maintain a CR or PR at Crossover Day 56. Durable ORR at Crossover Day 56 will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] based on the CAS using local investigators' review of aGvHD assessment data. Note that response is relative to the last assessment of aGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a durable responder at Crossover Day 56 if any of the following events occurs:

- Not a responder at Crossover Day 28
- Missing aGvHD assessment at Crossover Day 56
- No CR or PR at Crossover Day 56.
- Additional systemic therapy for aGvHD prior to Crossover Day 56

In addition, shift tables of aGvHD stage by organ and treatment group will be produced to compare Crossover Day 28 to Crossover Day 56 value for the patients who achieved PRs at Crossover Day 28.

2.7 Analysis of secondary efficacy objective(s)

The other secondary efficacy objectives are to:

- Evaluate the two treatment arms with respect to ORR at other time points, e.g. Day 14
- Evaluate the two treatment arms with respect to duration of response (DOR)
- Evaluate the two treatment arms with respect to overall survival (OS)
- Evaluate the two treatment arms with respect to event free survival (EFS)
- Evaluate the two treatment arms with respect to failure free survival (FFS)
- Evaluate the two treatment arms with respect to Non-relapse mortality (NRM)

- Evaluate the two treatment arms with respect to incidence of malignancy relapse/progression (MR)
- Describe cumulative steroid dosing until Day 56 in each treatment arm
- Evaluate the two treatment arms with respect to incidence of cGvHD
- Evaluate the two treatment arms with respect to BOR at any time points up to Day 28

All the secondary efficacy endpoint analyses are non-comparative in nature and will be analyzed using the Full Analysis Set (FAS).

2.7.1 Secondary efficacy endpoints

Overall Response Rate at Day 14

Overall Response Rate at Day 14 is defined as the proportion of patients with complete response (CR) or partial response (PR) at Day 14 according to standard criteria [Harris 2016].

Duration of response (DOR)

Duration of response is defined for patients whose overall response at Day 28 is complete response (CR) or partial response (PR) according to standard criteria [Harris 2016]. The start date is the date of first documented response of CR or PR (i.e., the start date of response), which could be prior to or at Day 28. If it's prior to Day 28, there should not be progression or addition of systemic therapies for aGvHD between the start date of response and Day 28. The end date is defined as the date of progression or the date of addition of systemic therapies for aGvHD on or after Day 28.

Death without prior observation of aGvHD progression and onset of chronic GvHD are considered to be competing risks.

Duration of response will be censored at the last response assessment prior to or at the analysis cut-off date, if no events/competing risk occurred before or at the cut-off date.

Overall survival (OS)

Overall survival is defined as the time from date of randomization to date of death due to any cause. If a patient is not known to have died, then OS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Event Free Survival (EFS)

Event-free survival is defined as the time from the date of randomization to the date of hematologic disease relapse/progression, graft failure or death due to any cause. If a patient is not known to have any event, then EFS will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Failure Free Survival (FFS)

Failure-free survival is defined as the time from the date of randomization to date of hematologic disease relapse/progression, non-relapse mortality or addition of new systemic aGvHD treatment.

Non-relapse mortality (NRM)

Non-relapse mortality is defined as the time from date of randomization to date of death not preceded by hematologic disease relapse/progression. Hematologic disease relapse/progression is considered a competing risk for NRM with the date of hematologic disease relapse/progression being the earlier of documented hematologic disease relapse/progression or institution of therapy to treat potential hematologic disease relapse/progression. If a patient is not known to have died or to have relapsed/progressed, then NRM will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Incidence of Malignancy Relapse/Progression (MR)

Malignancy relapse/progression is defined as the time from date of randomization to hematologic malignancy relapse/progression. Deaths not preceded by hematologic malignancy relapse/progression are competing risks. If a patient is not known to have event or competing risks, then MR will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Incidence of cGvHD

cGvHD is defined as the diagnosis of any cGvHD including mild, moderate, severe. Incidence of cGvHD is the time from date of randomization to onset of cGvHD. Deaths without prior onset of cGvHD and hematologic disease relapse/progression are competing risks. If a patient is not known to have event or competing risks, then the incidence of cGvHD will be censored at the latest date the patient was known to be alive (on or before the cut-off date).

Best Overall Response

Best overall response (BOR) is defined as proportion of patients who achieved overall response (CR or PR) at any time point up to and including Day 28, and had no additional systemic therapy for aGvHD prior to the time point.

2.7.2 Statistical hypothesis, model, and method of analysis

Overall Response Rate at Day 14

Overall response rate will be calculated based on the local investigators review of aGvHD assessment data. It will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

Duration of response (DOR)

The estimated cumulative incidence rates and 95% confidence intervals at 1, 2, 6, 12, 18 and 24 months will be presented for each treatment group. The cumulative incidence curve will be plotted. DOR will be listed and summarized by treatment group for all patients in the FAS with overall response of CR or PR at Day 28.

Overall survival (OS)

OS will be analyzed according to the randomized treatment group and strata assigned at randomization (aGvHD grade: Grade II vs. III vs. IV). The OS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 1, 2, 6, 12, 18 and 24

month survival estimates and 95% confidence intervals [Brookmeyer and Crowley 1982] will be presented for each treatment group. The hazard ratio for OS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

Event Free Survival (EFS)

EFS will be analyzed according to the randomized treatment group and strata assigned at randomization. The EFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 1, 2, 6, 12, 18 and 24 month survival estimates and 95% confidence intervals [Brookmeyer and Crowley 1982] will be presented for each treatment group. The hazard ratio for EFS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

A sensitivity analysis including aGvHD progression as an event will be performed using the same approach.

Failure Free Survival (FFS)

Cumulative incidence curve for FFS as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for each treatment group, accounting for onset of chronic GvHD as the competing risk.

In addition, the cumulative incidence of each of the three components considering the other two components as a competing risks will be estimated. Onset of chronic GvHD is considered as a competing risk for all three types of failure. The cumulative incidence curves will be plotted for each treatment group.

A sensitivity analysis including aGvHD progression as an event will be performed using the same approach.

Non-relapse mortality (NRM)

- NRM will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization (aGvHD grade: Grade II vs. III vs. IV). The cumulative incidence curve for NRM as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for each treatment group.
- As a sensitivity analysis, the cumulative incidence curve for NRM as well as estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for patients with underlying hematologic malignant disease in each treatment group.

Incidence of Malignancy Relapse/Progression (MR)

The cumulative incidence curve for MR and estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented for patients with underlying hematologic malignant disease, accounting for NRM as the competing risk.

In addition, the proportion of patients who had hematologic malignancy relapse/progression and its 95% confidence interval at 1, 2, 6, 12, 18 and 24 months will be presented by treatment group for patients with underlying hematologic malignant disease. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

Cumulative steroid dosing until Day 56

Overall and weekly cumulative steroid dose for each patient up to Day 56 or discontinuation of randomized treatment will be tabulated. In addition, the relative dose intensity (RDI), by week, will be calculated relative to the starting dose of corticosteroids and categorized as (1) complete reduction where patients are tapered off corticosteroids by D56, (2) ≤50% RDI and (3) >50% RDI. The proportion of patients in each category and corresponding 95% confidence intervals will be presented by treatment group. Average corticosteroid dose during the week ending on Days 14, 28, 56, 84, and 168 will also be tabulated and plotted.

Incidence of cGvHD

The cumulative incidence of cGvHD and estimates at 1, 2, 6, 12, 18 and 24 months with 95% confidence intervals will be presented, accounting for competing risks. The cumulative incidence curve will be plotted for each treatment group.

Best Overall Response

Best overall response will be calculated based on the local investigators review of aGvHD assessment data. It will be summarized using descriptive statistics (N, %) by treatment arm along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934]. Odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will be also presented.

2.7.3 Handling of missing values/censoring/discontinuations

Refer to Section 2.5.3

2.8 Safety analyses

All safety analyses will be based on Safety Set, except that the summary of safety data during the crossover treatment phase will be based on Crossover Analysis Set. All listings and tables will be presented by treatment group. For safety evaluations (except for AE) during randomized treatment phase, the last available assessment on or before the date of start of randomized treatment is taken as the “baseline” assessment. For safety evaluations (except for AE) during crossover treatment phase, the last available assessment on or before the date of start of crossover treatment is taken as the “baseline” assessment.

Due to possible crossover from BAT to ruxolitinib arm after Day 28, imbalance in exposure between the two arms is expected. Therefore, safety summaries for the randomized treatment will be performed for the following periods, unless specified:

- Up to Day 31 (the upper bound of the Day 28 visit window);
- Up to the earlier of i) cutoff date, ii) end date of on-randomized-treatment period;

The on-randomized-treatment period is defined in Section 2.1.1.

For the summaries up to cutoff date, some may be presented by adjusting for the duration of randomized treatment period in patient-years where relevant.

2.8.1 Adverse events (AEs)

For reporting of AEs the overall observation period will be divided into mutually exclusive categories, including pre-treatment, on-treatment (randomized or crossover), post-treatment periods as defined in Section 2.1.1.

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

AEs will be summarized by number and percentage of patients having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding. A patient with multiple occurrences of an AE will be counted only once in the respective AE category. A patient with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

In AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on the descending frequency in the ruxolitinib arm.

The following adverse event summaries will be produced by treatment arm; overview of adverse events and deaths, AEs by SOC and PT, summarized by relationship (all AEs and AEs related to study treatment), seriousness (SAEs and non-SAEs), leading to treatment discontinuation, leading to dose interruption/adjustment, and leading to fatal outcome. In addition, a summary of serious adverse events with number of occurrences will be produced (an occurrence is defined as >1 day between start and prior end date of record of same preferred term).

2.8.1.1 AEs adjusted for patient duration of treatment period

In order to account for differences in exposure of the ruxolitinib arm relative to the BAT arm due to crossover from BAT to ruxolitinib after Day 28 visit, incidence rates of adverse events may be presented by adjusting for duration of treatment period in patient-years where relevant.

2.8.1.2 Adverse events of special interest / grouping of AEs

Data analysis of AESIs

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to compound ruxolitinib. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and/or PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad.

For each AESI (selection from the Case Retrieval Sheet), number and percentage of patients with at least one event of the AESI occurring during on-treatment period will be summarized. Summaries of AESIs will be provided by treatment arm, (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose adjustment/interruption, hospitalization, death etc.).

A listing of all grouping levels down to the MedDRA preferred terms used to define each AESI will be generated.

In addition to summarizing infections by CTCAE grade, they will also be summarized using infection severity (protocol Appendix 2) up to Day 31 (the upper bound of the Day 28 visit window), and cutoff date.

Proportion of patients developing second primary malignancies will be summarized for on-treatment period, post-treatment period and both.

2.8.2 Deaths

Separate summaries for on-treatment and all deaths (including post-treatment death) will be produced by treatment arm, system organ class and preferred term.

All deaths will be listed and post-treatment deaths will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened patients.

2.8.3 Laboratory data

On analyzing laboratory, data from all sources (central and local laboratories) will be combined. The summaries will include all assessments available for the lab parameter collected no later than end of the on-treatment periods (randomized or crossover).

Grading of laboratory values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher.

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

- Shift tables using CTC grades to compare baseline to the worst on-treatment value

The following listings will be produced for the laboratory data:

- Listings of all laboratory data, with CTC grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.
- Listing of all CTC grade 3 or 4 laboratory toxicities

In addition to the above mentioned tables and listings, other exploratory analyses, for example figures plotting time course of raw or change in laboratory tests over time or box plots may be produced.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

Not applicable.

2.8.4.2 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: height (cm), weight (kg), body mass index (kg/m2), body temperature (°C), pulse (beats per minute), systolic and diastolic blood pressure (mmHg).

Data handling

Vital signs collected on treatment will be summarized. Values measured outside of on treatment period will be flagged in the listings.

Data analysis

For analysis of vital signs the clinically notable vital sign criteria are provided in Table 2-5 below.

Table 2-5 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Weight (kg)	increase > 10% from Baseline	decrease > 10% from Baseline
Systolic blood pressure (mmHg)	>=180 with increase from baseline of >=20	<=90 with decrease from baseline of >=20
Diastolic blood pressure (mmHg)	>=105 with increase from baseline of >=15	<=50 with decrease from baseline of >=15
Pulse rate (bpm)	>=100 with increase from baseline of >25%	<=50 with decrease from baseline of > 25%
Body temperature	>= 39.1	-

The number and percentage of patients with notable vital sign values (high/low) will be presented by treatment arm.

A listing of all vital sign assessments will be produced by treatment arm and notable values will be flagged. In the listing, the assessments collected outside of on-treatment period will be flagged.

Additional Analyses

Time to first occurrence of grade 3 infection

Time to first occurrence of infection is defined as time from start of study treatment to the date of first occurrence of grade 3 infection severity per protocol Appendix 2, i.e. time in days is calculated as (start date of first occurrence of infection) – (start of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be **the earliest** of the following dates:

- death date
- end date of on-treatment period (randomized or crossover)
- data cut-off date
- withdrawal of informed consent date

Failure curves (ascending Kaplan-Meier curves) will be constructed by treatment arm. Median together with 95% confidence interval as well as 25th percentile and 75th percentile will be presented for each treatment arm.

A sensitivity analysis will be conducted considering deaths or end of treatment phase (randomized or crossover) without prior infection as competing risks. Cumulative incidence curve for time to grade 3 infection as well as estimates at 1, 2 and 6 months with 95% confidence intervals will be presented for each treatment group.

In addition, the median time to occurrence for the subset of patients who experienced infection will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.

2.9 Pharmacokinetic endpoints

Pharmacokinetic Analysis Set (PAS) will be used in all pharmacokinetic data analysis and PK summary statistics.

Plasma samples for extensive PK will be taken at Day 1 (start of treatment), at Day 7 (week 1) for the first 25 adult patients and all adolescents to characterize the PK after first dose, and at steady state by non-compartmental analysis. Additional PK samples will be taken at later visits for all patients to characterize exposure-efficacy, exposure-safety [redacted] as data allows. Concentrations will be expressed in mass per volume units.

PK parameters

The PK parameters that will be determined are shown in Table 2-6. The PK parameters of ruxolitinib will be calculated from the extensive PK data based on the non-compartmental methods using Phoenix WinNonlin® (Pharsight, Mountain View, CA) software. Additional PK parameters may be estimated as needed.

Table 2-6 Non-compartmental PK parameters for ruxolitinib

AUClast	The AUC from time zero to the last measurable concentration sampling time (tlast) (ng x hr/mL)

AUCinf	The AUC from time zero to infinity (ng x hr/mL)
AUCtau	The AUC calculated to the end of a dosing interval (12 hr) at steady-state (ng x hr/mL)
Cmax	The maximum (peak) observed plasma, blood, serum, or other body fluid drug concentration after single dose administration (ng/mL)
Ctrough	The observed plasma concentration obtained prior to administration of the next dose (pre-dose concentration) (mass x volume-1)
Tmax	The time to reach maximum (peak) plasma, blood, serum, or other body fluid drug concentration after single dose administration (thr)
Lambda_z	Smallest (slowest) disposition (hybrid) rate constant (hr^{-1}) may also be used for terminal elimination rate constant (hr^{-1})
T1/2	The elimination half-life associated with the terminal slope (λz) of a semi logarithmic concentration-time curve (hr).
CL/F	The total body clearance of drug from the plasma (L/hr)
Vz/F	The apparent volume of distribution during terminal phase (associated with λz) (L)
Racc	Accumulation ratio (AUC at steady state/AUC Day 1)

Descriptive statistics (n, arithmetic mean, CV% mean, standard deviation (SD), median, geometric mean, CV% geo-mean, minimum and maximum) will be presented for the ruxolitinib arm for Pharmacokinetic Analysis Set for all PK parameters defined in Table 2-6 except Tmax, where only n, median, minimum and maximum will be presented.

All individual PK parameters will be listed for the ruxolitinib arm using the Full Analysis Set.

The potential impact of severity of GI GvHD on ruxolitinib pharmacokinetic parameters will be explored by summarizing PK parameters by baseline lower GI stage and producing box plots of AUCinf, AUCtau and Cmax by lower GI stage.

PK parameters will also be summarized by patients who took concomitant CYP3A4 inhibitors versus those who didn't. Patients are defined as having taken concomitant CYP3A4 inhibitors if based on the concomitant medication summary they have taken CYP3A4 inhibitors on the day of or on the day before PK samples are taken. In addition to the summary tables box plots of Cmax, AUCinf and AUCtau will be produced by concomitant CYP3A4 inhibitor use. This will be repeated for Day 1 and Day 7 PK parameters. The list of CYP3A4 inhibitors will be documented in Data Handling Plan.

[redacted]

[redacted]

PK concentrations

Descriptive statistics (n, m (number of non-zero concentrations), arithmetic mean, CV% mean, SD, median, geometric mean, CV% geo-mean, minimum and maximum) for ruxolitinib concentration by lower GI stage and overall will be presented at each scheduled time point for the ruxolitinib arm for the Pharmacokinetic Analysis Set.

Individual concentration-time profiles for ruxolitinib concentrations with median will be displayed graphically for the ruxolitinib arm for Full Analysis Set on the semi-log view. In addition, the mean (+/- SD) and geometric mean concentration-time profiles for ruxolitinib by treatment over time will be displayed graphically for Pharmacokinetic Analysis Set on the linear and semi-log view.

All individual plasma ruxolitinib concentration data will be listed for the ruxolitinib arm for the Full Analysis Set.

Crossover concentrations will be included in all analyses on concentrations.

Handling of PK data below LLOQ or missing

All concentration values below the lower limit of quantitation (LLOQ) are set to zero by the Bioanalyst, and will be displayed in the listings as zero and flagged. LLOQ values will be treated as zero in any calculations of summary statistics, and treated as missing for the calculation of the geometric means and their CV%. The number of non-zero concentrations will also be reported in the summary statistics.

Missing values for any PK data will not be imputed and will be treated as missing.

2.10 PD and PK/PD analyses

Analysis of relationship between ruxolitinib exposure and efficacy/safety endpoints

This analysis will be described and reported separately to the CSR. The following are objectives for exposure-response analysis:

- Characterize the exposure-efficacy relationship of ruxolitinib in terms of exposure and efficacy response, with efficacy response defined as(overall response rate at Day 28 and durable response at Day 56, duration of response, overall survival at 6 months and any other relevant endpoints, and exposure defined plasmaconcentration, PK parameter or dose, as appropriate)
- Characterize the exposure-safety relationship of ruxolitinib in terms of exposure and safety response, where safety response is defined as various categorizations of AEs

(frequency of;severity of AEs, AEs of interest to be defined prior to analysis or laboratory parameters).

- Average steady-state exposures and/or other PK parameters for the population will becomputed by the POP PK model accounting for dose modifications or dose interruptions up to the day prior to the day of assessments. Population PK derived parameters will be used for exposure-response analysis by appropriate methods

For duration of response and overall survival, if the primary endpoint is significant and sufficient events have accrued (i.e., at analysis time points after the primary analysis), a Cox regression model with appropriate patient demographic and prognostic factors as covariates and the log-trough level as a time dependent covariate will be fitted if appropriate. To account for dose adjustments between trough sampling time points the trough concentration will be adjusted by dividing by the actual dose prior to the trough in question and multiplying by the arithmetic mean dose since the last trough sample or start of dosing as applicable. The calculation of the arithmetic mean dose should count any dose interruptions as a zero dose for the days that no dose was given. Goodness of fit of the model will also be examined. The survival time should be calculated from the time of first dose rather than the time of randomization. The hazard ratio and 95% confidence interval for a two-fold increase in exposure will be displayed if appropriate. Kaplan-Meier curves may be used to summarize the data based on relevant quantiles of PK.

For incidence of specific AEs, Day 28 response (if the primary endpoint is significant) and Day 56 durable response (if the primary and key secondary endpoints are significant), logistic regression models may be used including log-average trough concentration and other demographic and prognostic covariates in the model as appropriate. The trough concentration will be similarly adjusted to account for dose adjustments as described above. The average trough concentration included in the model for Day 28 response will be the arithmetic mean of the Day 7, Day 14, Day 21 and Day 28 adjusted trough concentration. A similar average trough concentration will be calculated for the Day 56 response except the Day 42 and Day 56 adjusted concentration will be double weighted compared to the others because they are representative of concentration over 2 weeks rather than 1. For analysis of adverse events the last trough concentration prior to the AE will be included in the model. Other exposure measures may be considered if appropriate. The odds ratio and its 95% confidence interval for a two-fold increase in exposure will be displayed as appropriate.

2.11 Patient-reported outcomes

The FACT-BMT along with the EQ-5D-5L will be used to collect data on the patient's disease-related symptoms and health-related quality of life. Responses to the FACT-BMT and EQ-5D-5L will be generated in accordance with the respective scoring manual. Further details on the scoring of FACT-BMT and EQ-5D are given in Appendix 5.4.3.1 and Appendix 5.4.3.2.

These PRO instruments are planned to be administered on randomization day and every week during the first 2 months, and every 4 weeks thereafter until the end of treatment.

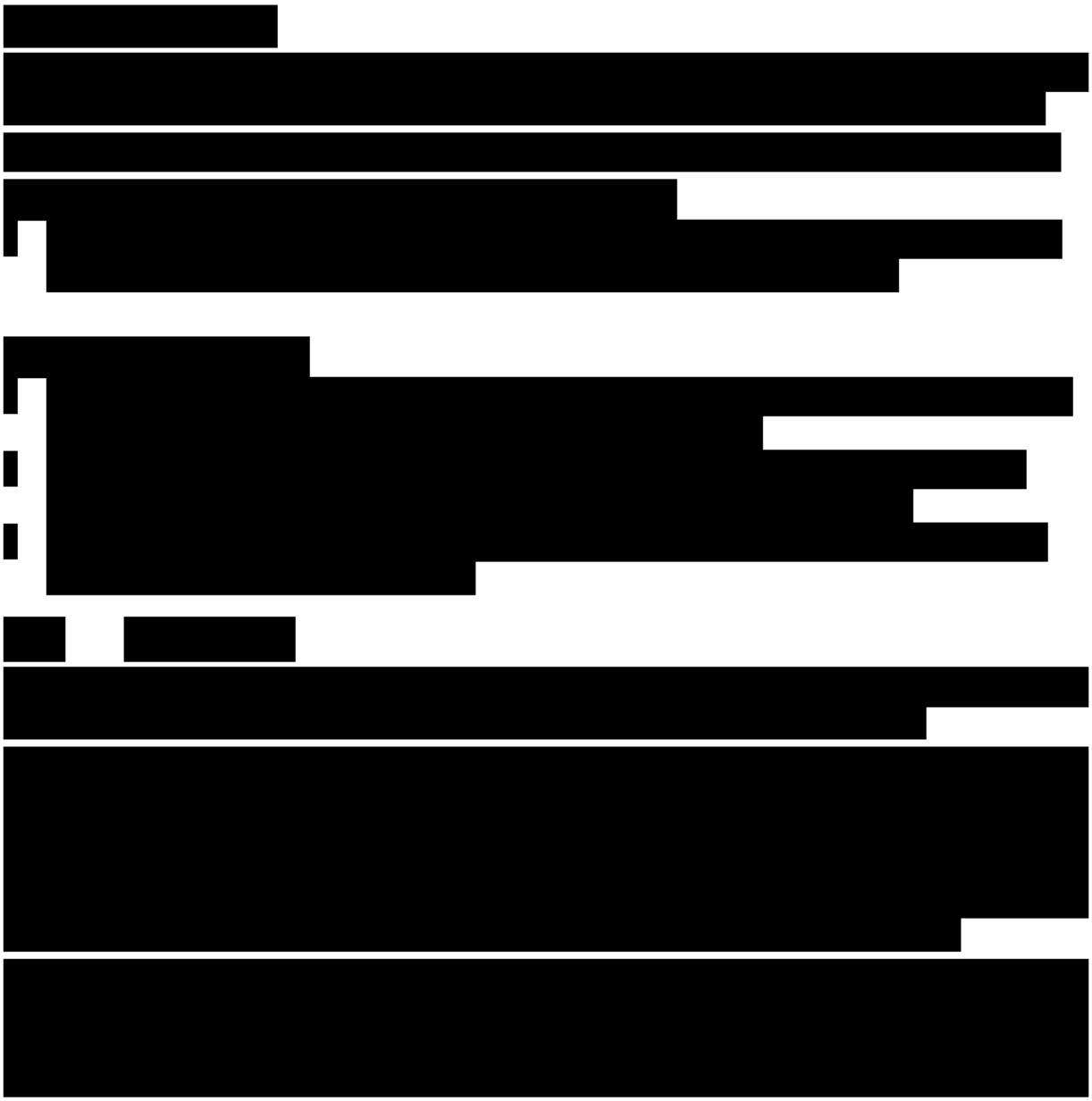
The baseline is defined as the last PRO assessment prior to or on the date of randomization date + 3 days, but no later than the treatment start date..

FACT-BMT will not be used in patients under age 18 years.

Descriptive statistics (mean, standard deviation, median, Q1, Q3, minimum, and maximum) will be used to summarize the scored scales at each scheduled assessment time point for the FACT-BMT and EQ-5D using FAS. Additionally, change from baseline in the scores at the time of each assessment will be summarized. Patients with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

Missing items data in a scale will be handled based on each instrument manual. No imputation will be applied if the total or subscale scores are missing at a visit.

Subscale scores from the FACT-BMT and EQ-5D-5L will be displayed as mean profiles for each study arm, presented over time using time windows as described in Section 2.1.



[redacted]

[redacted]

2.14 Interim analysis

No formal interim analysis is planned for this trial.

Early PK analysis

To compare the exposure of 10 mg BID in the SR aGvHD population to the known exposure in MPN patients, the early extensive ruxolitinib PK data on the first 25 adult patients (including any adolescent patients randomized at that time) will be explored once available, also in the context of concomitant medications. However, there will be no comparison between the two treatment arms because PK data are not collected in BAT patients.

DMC safety data analysis

DMC will be instituted in this study and will review the safety data as outlined according the DMC charter. A separated set of SAP and TFL shells will be provided to detail the analysis.

3 Sample size calculation

The study with a total of 308 patients and 1:1 randomization (ruxolitinib vs. BAT) stratified on aGvHD grade (Grade II vs. Grade III vs. Grade IV) has 90% power to test for the primary endpoint (ORR at Day 28) and approximately 90% power to test for the key secondary endpoint (durable ORR at Day 56). The family wise α-level will be controlled at 0.025 overall for the two comparisons. Specifically, this study will claim to have achieved the efficacy objective when the primary endpoint ORR at Day 28 shows a significant treatment effect at one-sided $\alpha = 0.025$. Conditional on significance of the primary endpoint, the key secondary endpoint durable ORR at Day 56 will be tested at one-sided $\alpha = 0.025$.

3.1 Primary analysis

The sample size calculation is based on the primary variable ORR at Day 28. The hypotheses to be tested and details of the testing strategy are described in Section 2.5.2 and Section 3.

Based on Martin P. et al BBMT 2013, the ORR at Day 28 in the BAT arm is expected to be 58%. The stratum specific rates (Grade II 69%, Grade III 59%, Grade IV 50%) are obtained assuming that the ratio of aGvHD Grade II: III: IV is 0.2:0.4:0.4. It is expected that treatment with ruxolitinib will result in an 18% increase in the ORR, i.e., an expected odds ratio of 2.25 (which corresponds to an increase in ORR to 75%). Power for the CMH test, stratifying on aGvHD grade, was calculated using software package East V6. In order to ensure 90% power a total sample size of 308 patients are needed.

With a sample size of 154 patients in each treatment arm, an observed odds ratio greater than or equal to 1.63 would achieve statistical significance for the primary endpoint. Assuming that the observed response rates in Grades II/III/IV in BAT arm are 69%/59%/50% (overall 57%), observed response rates ≥ 78%/70%/62% (overall 68%) in the ruxolitinib arm would achieve statistical significance.

3.2 Power for analysis of key secondary variables

Durable ORR at Day 56, as the key secondary variable, will be formally statistically tested, provided that the primary endpoint ORR at Day 28 is statistically significant. The hypotheses

to be tested and details of the testing strategy are provided in Section 2.6.2. Based on [Van Groningen 2016], the durable ORR at Day 56 in the BAT arm is expected to be approximately 35%. The stratum specific rates (Grade II 45%, Grade III 36%, Grade IV 30%) are obtained assuming that the ratio of aGvHD Grade II: III: IV is 0.2:0.4:0.4. It is expected that treatment with ruxolitinib will result in a 20% increase in the ORR, i.e., an expected odds ratio of 2.25 (which corresponds to an increase in durable ORR to 55%). With these assumptions and sample size of 308 patients, the power for the key secondary endpoint is at least 90%.

With sample size 154 patients in each treatment arm, an observed odds ratio greater than or equal to 1.59 for durable ORR at Day 56 would achieve statistical significance. Assuming that the observed durable response rates in Grades II/III/IV in BAT arm are 45%/36%/30% (overall 35%), observed durable response rates ≥ 57%/47%/41% (overall 47%) in the ruxolitinib arm would achieve statistical significance.

4 Change to protocol specified analyses

Compared to the protocol version 00, 'discontinuation from randomized treatment' has been dropped as reason. Since some BATs have fixed duration which is less than 8 weeks, to avoid bias, discontinuation from randomized treatment will not be used to disqualify the response or durable response. Also duration of response will not be censored based on treatment discontinuation.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

The following rule should be used for the imputation of the dose end date for a given study treatment component:

Scenario 1: If the dose end date is completely missing and there is no EOT page and no death date, the patient is considered as on-going:

The patient should be treated as on-going and the cut-off date should be used as the dose end date.

Scenario 1 should not be applied for final CSR. All patients should have EOT page complete before the Database lock for Final CSR

Scenario 2: If the dose end date is completely or partially missing and the EOT page is available:

Please note that date of assessment on EOT eCRF might be very different from last date of dose.

Case 1: The dose end date is completely missing, and the EOT completion date is complete, then this latter date should be used.

Case 2: Only Year(yyyy) of the dose end date is available and yyyy < the year of EOT date:

Use Dec31yyyy

Case 3: Only Year(yyyy) of the dose end date is available and yyyy = the year of EOT date:

Use EOT date

Case 4: Both Year(yyyy) and Month (mm) are available for dose end date, and yyyy = the year of EOT date and mm < the month of EOT date:

Use last day of the Month (mm)

All other cases should be considered as a data issue and the statistician should contact the data manager of the study.

After imputation, compare the imputed date with start date of treatment, if the imputed date is < start date of treatment:

Use the treatment start date

Patients with missing start dates are to be considered missing for all study treatment component related calculations and no imputation will be made. If start date is missing then end-date should not be imputed.

5.1.2 AE, ConMeds and safety assessment date imputation

Table 5-1 Imputation of start dates (AE, CM) and assessments (LB, EG, VS)

Missing Element	Rule
day, month, and year	• No imputation will be done for completely missing dates

day, month	• If available year = year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY
	  ○ Else set start date = study treatment start date.
	• If available year > year of study treatment start date then 01JanYYYY
	• If available year < year of study treatment start date then 01JulYYYY

Missing Element	Rule
day	• If available month and year = month and year of study treatment start date then o If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY. o Else set start date = study treatment start date. • If available month and year > month and year of study treatment start date then 01MONYYYY • If available month and year < month year of study treatment start date then 15MONYYYY

Table 5-2 Imputation of end dates (AE, CM)

Missing Element	Rule (* if end date of the on-treatment period not > (death date, cut-off date, withdrawal of consent date))
day, month, and year	• Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	• If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period *
day	• If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

Any AEs and ConMeds with partial/missing dates will be displayed as such in the data listings.

Any AEs and ConMeds which are continuing as per data cut-off will be shown as 'ongoing' rather than the end date provided.

The above imputations are only used for analyses of time to and duration of AEs and concomitant medications.

5.1.2.1 Other imputations

Incomplete date of initial diagnosis of aGvHD

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

5.2 AEs coding/grading

Adverse events are coded using the Medical dictionary for regulatory activities (MedDRA) terminology.

The latest available MedDRA version at the time of the analyses should be used. The MedDRA version used should be specified in the footnote of relevant tables.

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 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).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTCAE grades are given in Novartis internal criteria for CTCAE grading of laboratory parameters. The latest available version of the document based on the underlying CTCAE version 4.03 at the time of analysis will be used. For laboratory tests where grades are not defined by CTCAE v4.03, results will be graded by the low/normal/high (or other project-specific ranges, if more suitable) classifications based on laboratory normal ranges.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. CTCAE Grade 5 is not defined for laboratory values. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

Imputation Rules

CTC grading for blood differentials is based on absolute values. However, this data may not be reported as absolute counts but rather as percentage of WBC.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

The following rules will be applied to derive the WBC differential counts when only percentages are available for a xxx differential

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

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4]

In order to apply the above formula, albumin values in g/L will be converted to g/dL by multiplying by 0.1), calcium values in mmol/L will be converted to mg/dL by dividing by 0.2495. For calculation of laboratory CTC grades 0 and 1, the normal range for derived corrected calcium is set to the same limits (in mg/dL) as for calcium.

CTC grades for the derived absolute WBC differential counts (neutrophils, lymphocytes) and corrected calcium will be assigned as described above for grading.

5.4 Statistical models

5.4.1 Primary analysis

The null hypothesis of equality of response rate in the two treatment arm will be tested against one-sided alternative. The statistical hypotheses are:

$$H_0: ORR_{rux} \leq ORR_{BAT} \quad vs. \quad H_1: ORR_{rux} > ORR_{BAT}$$

where ORR_{rux} is the probability of response in ruxolitinib and ORR_{BAT} is the probability of response in BAT.

The Cochran-Mantel-Haenszel chi-square test X^2_{CMH} (implemented again via SAS procedure FREQ with CMH option in the TABLES statement) will be used to test the difference in response rates between the treatment arms. The p-value corresponding to the CMH test for "general association" will be used which follows a Chi-square distribution with one degree of freedom.

If the sampling assumptions for chi-square test is not met, the exact Cochran-Mantel-Haenszel test will be used (implemented via SAS procedure MULTTEST). The test is performed by running a stratified version of the Cochran-Armitage permutation test [Armitage et al. 1969]. In studies with stratified randomization, the chi-square approximation is considered appropriate for the X^2_{CMH} statistics if the rule of Mantel and Fleiss [Mantel and Fleiss 1980] is satisfied.

Logistic Regression

Odds ratio will be used as a measure of association between treatment and response. The odds ratio will be derived from the logistic regression model (implemented using SAS procedure LOGISTIC, with treatment specified as an explanatory variable in the CLASS statement) which allows for including not only the stratification factor but also for adjustments for other covariates (both categorical and continuous). The odds ratio will be presented with 95% Wald confidence limits.

In cases where an exact test has been used to compare response rates, the odds ratio should be determined using exact logistic regression, and the odds ratio presented with exact 95% confidence limits. In these cases, SAS PROC LOGISTIC with EXACTONLY option will be used.

Confidence interval for response rate

Responses will be summarized in terms of percentage rates with $100(1-\alpha)\%$ confidence interval using exact binomial confidence interval (implemented using SAS procedure FREQ with EXACT statement for one-way table [Clopper and Pearson 1934].

Definition of new or additional systemic therapy for aGvHD

The data source to search for the new or additional systemic therapy would be from eCRFs of Prior and Concomitant Medication (CONMED) (with subcategory as treatment for aGvHD) and Dosage Administration Record (DAR).

Any of the following therapies represent new or additional systemic therapy for aGvHD:

1. Any new CNI therapy being initiated as 'treatment for aGvHD' after the baseline as recorded either on DAR or on CONMED, and never received prior to or at baseline
2. Any other systemic therapy (excluding CNI or systemic corticosteroid (methylprednisolone, prednisone and prednisolone)) being started as 'treatment for aGvHD' after the baseline and recorded on CONMED. Note: the therapy may have been taken prior to baseline but not at baseline.
3. Additional BAT at or after start of the initial BAT as recorded on DAR
4. Treatment with ruxolitinib after cross-over from BAT as recorded on DAR

5.4.2 Key secondary analysis

Same instructions as in Section 5.4.1.

5.4.3 Other secondary analysis

5.4.3.1 FACT-BMT

For FACT-BMT the subscale scores, the FACT-BMT total score and the FACT-BMT Trial Outcome Index (TOI) will be calculated.

The subscales are physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), functional well-being (FWB) and bone marrow transplant subscale (BMTS). For each subscale the corresponding score will be calculated based on the item response of the answered question according to the FACT-BMT Scoring Guide (Version 4). The scoring guide identifies those items that must be reversed before being added to obtain subscale totals. Negatively stated items are reversed by subtracting the response from "4". After reversing proper items, all subscale items are summed to a total, which is the subscale score.

If there are missing items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done by using the formula below:

Prorated subscale score

= [Sum of item scores] x [N of items in subscale]/[N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the relevant items were answered. Note that not all items (questions) are used to calculate the scores (see Scoring Guide, version 4). For all the scores the higher the score is the better the QOL is.

The FACT-BMT Trial Outcome Index (TOI, score range: 0-96) is calculated as

$$\underset{(\textbf{PWB score})}{\_\_\_\_\_} + \underset{(\textbf{FWB score})}{\_\_\_\_\_} + \underset{(\textbf{BMTS score})}{\_\_\_\_\_} = \_\_\_\_\_ = \textbf{FACT-BMT TOI}$$

The FACT-BMT total score (score range: 0-148) is then calculated as the sum of all un-weighted subscale scores:

(**PWB** score) + (**SWB** score) + (**EWB** score) + (**FWB** score) + (**BMTS** score) = **FACT-BMT Total score**

5.4.3.2 EQ-5D-5L

EQ-5D-5L consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Also the EQ-VAS is included. Each dimension has 5 levels (1-5).

The data were reported for all 5 dimensions. The proportion of reported problems for each level were reported. The changes from baseline were defined with a shift table. The below screenshot is from the Excel conversion document obtained from the EuroQol webpage on 5L Value Sets (http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html). This UK value will be used to convert the "health state" to an "index value" for all patients in all countries. For example, if the health state is 11125, then the index value would be .316 and so on.

EQ-5D-5L_Crosswalk_Value_Sets.xls [Read-Only] [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J	K	L
1		Health state	Denmark	France	Germany	Japan	Netherlands	Spain	Thailand	UK	US	Zimbabwe
2		5L profile	Denma	Franc	Germa	Jap	Netherlan	Spa	Thaila	U	U	Zimbab
3		11111	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.900
4		11112	0.856	0.929	0.999	0.829	0.845	0.932	0.814	0.879	0.876	0.864
5		11113	0.818	0.910	0.999	0.785	0.805	0.914	0.766	0.848	0.844	0.854
6		11114	0.671	0.769	0.809	0.761	0.592	0.731	0.660	0.635	0.700	0.792
7		11115	0.519	0.622	0.611	0.736	0.370	0.541	0.549	0.414	0.550	0.727
8		11121	0.859	0.910	0.910	0.814	0.874	0.910	0.780	0.837	0.861	0.846
9		11122	0.787	0.839	0.909	0.740	0.765	0.857	0.723	0.768	0.820	0.810
10		11123	0.768	0.820	0.909	0.721	0.736	0.843	0.708	0.750	0.809	0.800
11		11124	0.622	0.679	0.719	0.697	0.523	0.660	0.602	0.537	0.669	0.738
12		11125	0.469	0.532	0.521	0.672	0.301	0.470	0.491	0.316	0.524	0.673
13		11131	0.824	0.888	0.887	0.768	0.843	0.887	0.726	0.796	0.827	0.833
14		11132	0.770	0.817	0.887	0.718	0.745	0.838	0.701	0.740	0.806	0.797
15		11133	0.756	0.798	0.887	0.705	0.719	0.825	0.694	0.725	0.800	0.787
16		11134	0.609	0.657	0.697	0.681	0.506	0.642	0.588	0.512	0.661	0.725
17		11135	0.457	0.510	0.499	0.656	0.284	0.452	0.477	0.291	0.517	0.660

6 Reference

1. Anderson JR, Cain KC and Gelber RD (1983). Analysis of survival by tumor response. Journal of Clinical Oncology, 1(11), 710-719.
2. Armitage P, McPherson CK and Rowe BC (1969). Repeated significance tests on accumulating data. J .R. Statist. Soc. A, 132, 232-244.
3. Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41.
4. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrical, 26, 404-413.
5. FDA briefing Document Oncologic Drugs Advisory Committee Meeting, July 24, 2012, Evaluation of radiologic review of progression-free survival in Non-hematologic malignancies. http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/oncologicdrugsadvisorycommittee/ucm311141.pdf
6. Harris AC, Young R, Devine S, Hogan WJ, Ayuk F, Bunworasate U, Chanswangphuwana C, Efebera YA, Holler E, Litzow M, Ordemann R, Qayed M, Renteria AS, Reshef R, Wölfl M, Chen YB, Goldstein S, Jagasia M, Locatelli F, Mielke S, Porter D, Schechter T, Shekhovtsova Z, Ferrara JL, Levine JE. (2016) International, Multicenter Standardization of Acute Graft-versus-Host Disease Clinical Data Collection: A Report from the Mount Sinai Acute GVHD International Consortium. Biol Blood Marrow Transplant; 22(1): 4-10.
7. Korhonen P, Laird NM and Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC study (1999). Statistics in Medicine, 18, 2879-2897.
8. Mantel, N. and Joseph L Fleisś (1980). Minimum expected cell size requirements for the Mantel-Haenszel one degree-of-freedom Chi-Square Test and a related rapid procedure. American Journal of Epidemiology, 112, 129-134.
9. Robins JM and Tsiatis AA (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models (RPSFTM). Commun. Statist.-Theory Meth., 20(8), 2609-31.
10. Van Groningen, LF, Liefferink AM, de Haan AF, Schaap NP, Donnelly JP, Blijlevens NM, van der Velden WJ. (2016) Combination Therapy with Inolimomab and Etanercept for Severe Steroid-Refractory Acute Graft-versus-Host Disease Biol Blood Marrow Transplant. 22(1):179-82.