Official Title:	**A Phase III Randomized Open-label Multi-center Study of Ruxolitinib vs. Best Available Therapy in Patients With Corticosteroid-refractory Chronic Graft vs Host Disease After Allogeneic Stem Cell Transplantation (REACH3)**
NCT Number:	NCT03112603
Document Date:	Statistical Analysis Plan: 06 July 2020

NOVARTIS

Incyte

Clinical Development

INC424/ruxolitinib/Jakavi®/Jakafi®

CINC424D2301 (INCB 18424-365)

A phase III randomized open-label multi-center study of ruxolitinib vs. best available therapy in patients with corticosteroid-refractory chronic graft vs host disease after allogenic stem cell transplantation (REACH 3)

Statistical Analysis Plan (SAP) – Amendment 2

TEMPLATE VERSION AND DATE: Final 2.0 22-Jul-2016, Effective 22-Jul-2016

Author: [redacted] (Trial Statistician)

Document type: SAP Documentation

Document status: Final

Release date: 06-Jul-2020

Number of pages: 54

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
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
eCRF	Electronic Case Report Form
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
[redacted]	[redacted]
PK	Pharmacokinetics
PPS	Per-Protocol Set
PR	Partial response
PRO	Patient-reported Outcomes
QoL	Quality of Life
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 CINC424D2301, a phase III, randomized, open-label, multi-center study of ruxolitinib versus best available therapy in patients with corticosteroid-refractory chronic graft vs. host disease after allogeneic stem cell transplantation.

This SAP is an amendment of SAP amendment 1. The main purpose of this amended SAP is to provide details on the overall hierarchical testing procedure for the primary and the two key secondary endpoints that was specified for the second interim analysis and documented in the DMC SAP. Furthermore, the definition of the Per-Protocol set was modified by adding further criteria to exclude patients for a sensitivity analysis. Finally, a few minor modifications and formal corrections were made.

1.1 Study design

This randomized, Phase III, open-label, multi-center study will investigate the efficacy and safety of ruxolitinib vs. Best Available Therapy (BAT), added to the subject's immunosuppressive regimen of corticosteroids ± calcineurin inhibitor (CNI) in adults and adolescents (≥12 years old) with corticosteroid-refractory chronic Graft vs Host Disease (SR-cGvHD).

This is an open-label study. Investigators, patients, and Sponsor will have full knowledge of the treatment allocation. Approximately 324 patients will be randomized in this study. Each patient will be treated and followed for a total of 3 years (39 cycles/156 weeks). The study design is illustrated in Figure 1-1 below, with each cycle comprised of 4 weeks (28 days).

Patients will be randomized in a ratio of 1:1 (ruxolitinib or BAT), randomization will be stratified by cGvHD severity: moderate vs severe (see Protocol Appendix 2).

The following periods were defined:

- **Primary Efficacy Period (Cycle 1 through end of Cycle 6)**

  Study treatment will begin on Cycle 1 Day 1 following randomization.

  Study visits will occur per the following schedule to monitor tolerability and efficacy of the study treatments during the Primary Efficacy Period:

  - Cycle 1: Day 1, 8, 15, 22
  - Cycle 2 to 6: Day 1 of each cycle

  Patients will be treated for a minimum of 6 cycles, until Cycle 7 Day 1, unless they experience intolerable toxicity, cGvHD progression, or withdraw from the study (see Protocol Section 6 for Study Treatment details and permitted/prohibited concomitant therapies).

  Addition or initiation of a new systemic therapy is allowed only after documented lack of response (as defined in Protocol Section 3), intolerable toxicity, or cGvHD flare and will be considered a treatment failure for both the primary and key secondary objectives. At the study visit in which the patient meets the criteria for disease progression, intolerable toxicity, or cGvHD flare treatment failure, addition or initiation of a new systemic

treatment is allowed. However, if mixed response or no response is assessed, addition or initiation of a new systemic treatment is allowed after confirmation at the next scheduled study visit, at least 4 weeks later (See Protocol Figure 4-2).

Primary efficacy assessments will be performed on Cycle 7 Day 1, after completion of 6 cycles of treatment.

- **Extension Period (Cycle 7 to Cycle 39)**

  After the Cycle 7 Day 1 primary efficacy assessment visit, study visits will occur every 3 cycles (every 12 weeks +/- 7 days) starting with Cycle 9 Day 1 and up to Cycle 39, or EOT, whichever occurs first (i.e. C9D1, C12D1, C15D1, etc.).

  Patients in the Extension Period will either continue the current BAT treatment or ruxolitinib treatment, cross over to the ruxolitinib treatment arm, or permanently discontinue study treatment and enter the Long-Term Survival Follow-Up. Addition or initiation of another BAT is not allowed in the Extension Period (See Protocol Figure 4-3).

- **Cross-Over for BAT patients only (from Cross-Over up to Cycle 33)**

  On Cycle 7 Day 1 and thereafter, patients randomized to BAT may cross over to ruxolitinib treatment due to disease progression, mixed response, or unchanged response, due to toxicity to BAT, or due to cGvHD flare. Cross-Over patients will switch to and follow the visit evaluation schedule as specified in Protocol Table 7-2. Patients crossing over from BAT will be treated and/or followed on study for a total of 39 cycles, inclusive of randomized treatment, Cross-Over treatment (BAT patients only), and long-term survival follow up (i.e. if patients cross over on randomized treatment Cycle 7, they will only complete up to cycle 33 in the cross over treatment period).

  Patients who meet Cross-Over criteria to receive ruxolitinib are allowed to continue their systemic immunosuppressive regimen of corticosteroids +/- CNI for SR-cGvHD treatment as per standard of care, with cessation required of the BAT treatment before starting treatment with ruxolitinib.

  Patients undergoing dose tapering following response should be monitored for cGvHD flare occurrences, which must be reported in the database.

- **Long-Term Survival Follow-Up (EOT to 39 cycles on study)**

  Patients who permanently discontinue the study treatment prior to completion of 39 cycles on study for reasons other than achieving a CR or PR will enter the Long-Term Survival Follow-Up (LTSFU), and may be treated per Institutional practice. They will be followed approximately every 3 months by telephone call for survival and reporting of new cGvHD therapies until 39 cycles are completed.

  New cGvHD therapies will be entered in the New cGvHD Treatment Since Discontinuation of Study Treatment eCRF.

  See Protocol Section 7.1.5 for permanent treatment discontinuation details.

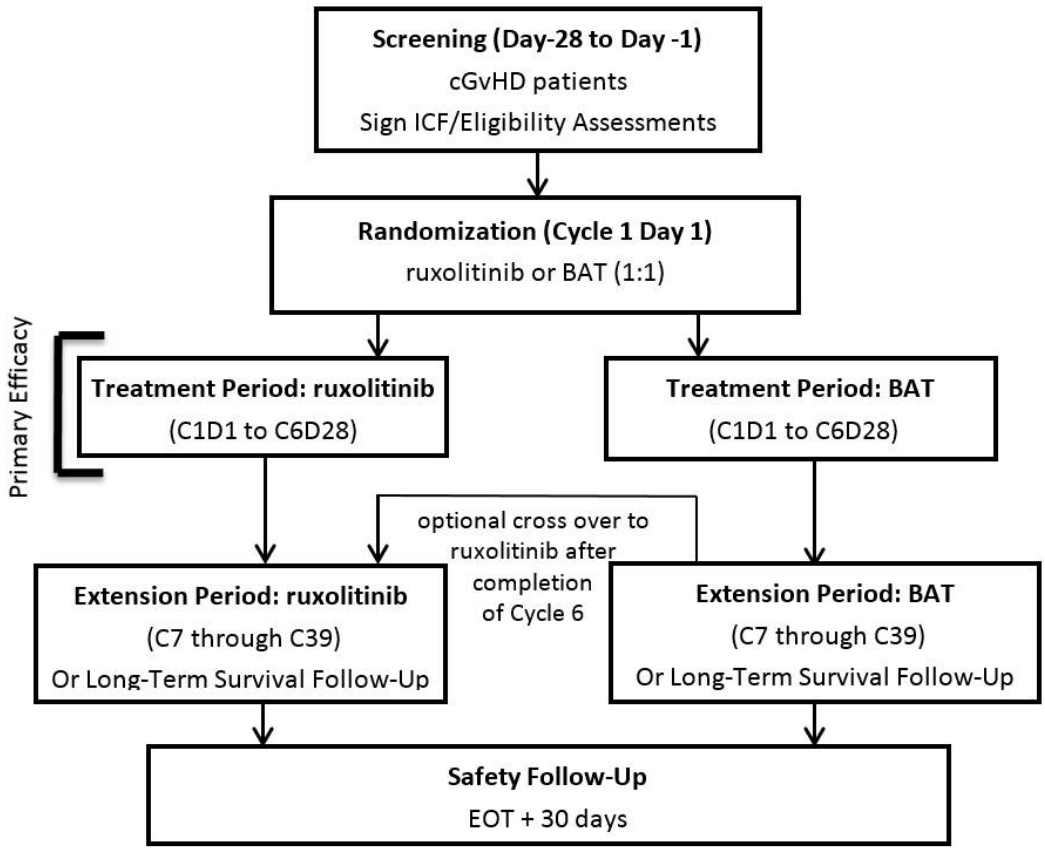
- **Safety Follow-Up (Last Dose + 30 days)**

  A 30-day safety follow-up assessment will be performed after the last dose of ruxolitinib or BAT, for patients that permanently discontinue study treatment for reasons other than CR or PR.

The primary efficacy variable, ORR at Cycle 7 Day 1, will be analyzed at the time when all patients have completed the Cycle 7 Day 1 visit or discontinued earlier. At this time, the primary clinical study report (CSR) will be produced. Following the cut-off date for the analysis reported in the primary CSR, the study will remain open. Ongoing patients will continue to receive study treatment and be followed as per the Visit Evaluation Schedule, as long as patients derive benefit from ruxolitinib or BAT, until completion of 39 cycles (156 weeks/3 years) of study treatment and/or follow up, inclusive of randomized treatment, cross over treatment (BAT patients only), and long-term survival follow up. The end of study is defined as the earliest occurrence of one of the following:

- All patients have completed 39 cycles OR discontinued from the study OR died

Figure 1-1 Schematic study design



1.2 Study objectives and endpoints

Table 1-1 Objectives and related endpoints

Objective	Endpoint	Analysis
Primary endpoint		Refer to Section 2.5
To compare the efficacy of ruxolitinib vs. Investigator's choice Best Available Therapy (BAT) in patients with moderate or severe SR-cGvHD assessed by Overall Response Rate (ORR) at the Cycle 7 Day 1 visit	Overall response rate (ORR) on Cycle 7 Day 1 after randomization, defined as the proportion of patients in each arm demonstrating a complete response (CR) or partial response (PR) without the requirement of additional systemic therapies for an earlier progression, mixed response or non-response. Scoring of response will be relative to the organ score at randomization.	CMH test, stratified by cGvHD severity (moderate vs. severe)
Key secondary endpoints		Refer to Section 2.6
To compare the rate of failure free survival (FFS) To compare change in modified Lee Symptom Score FFS will be used as the first key secondary endpoint for all regions except the US (ROW). The modified Lee symptom score will be used as the first key secondary endpoint for the US	Composite time to event endpoint incorporating the following FFS events: i) relapse or recurrence of underlying disease or death due to underlying disease, ii) non-relapse mortality, or iii) addition or initiation of another systemic therapy for cGvHD. Rate of patients with clinically relevant improvement of the modified Lee symptoms score at Cycle 7 Day 1	A fixed sequence hierarchical testing strategy will be applied for the primary and the two key secondary endpoints. For ROW FFS will be tested first followed by the modified Lee symptom score. For US, modified Lee symptom score will be tested first followed by FFS. FFS: Logrank-test, stratified by cGvHD severity (moderate vs. severe) Modified Lee symptom score: CMH test comparing rates of responders, stratified by cGvHD severity (moderate vs. severe)

Objective	Endpoint	Analysis
Other secondary endpoints		Refer to Section 2.7 Analysis of secondary efficacy objective(s), 2.8 Safety analyses and 2.9 Pharmacokinetic endpoints
Best overall response (BOR)	Proportion of patients who achieved OR (CR+PR) at any time point (up Cycle 7 day 1 or the start of additional systemic therapy for cGvHD)	
To estimate ORR at end of Cycle 3	Proportion of patients who achieved OR (CR+PR) at Cycle 4 Day 1.	
Duration of Response	Duration of response (DOR) is assessed for responders only. DOR is defined as the time from first response until cGvHD progression, death, or the date of change/addition of systemic therapies for cGvHD.	
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.	
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 underlying disease relapse/recurrence.	
To assess proportion of patients with ≥50% reduction in daily corticosteroid dose at Cycle 7 Day1		
To assess proportion of patients successfully tapered off all corticosteroids at Cycle 7 Day 1		
To assess cumulative incidence of Malignancy Relapse/Recurrence (MR)	Malignancy Relapse/Recurrence (MR) is defined as the time from date of randomization to hematologic malignancy relapse/recurrence. Calculated for patients with underlying hematologic malignant disease.	
To evaluate changes in FACT-BMT and EQ-5D	Change in FACT-BMT from baseline to each visit where measured. Change in EQ-5D from baseline to each visit where	

Objective	Endpoint	Analysis
	measured.	
To assess Pharmacokinetics (PK) of ruxolitinib in SR-cGvHD patients	Pharmacokinetic parameters of ruxolitinib after a single dose and at steady state. Cmax, AUClast, and AUCinf. Other PK parameters are CL/F, Vz/F, Tmax and T1/2.	
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.	
To assess medical resource utilization	Resources including duration and frequency of hospitalization, emergency room visits, additional outpatient office visits to general practitioner, specialist, and urgent care visits. Frequency of concomitant treatments will also be captured.	

[redacted]

Objective	Endpoint	Analysis
[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 the Cycle 7 Day 1 visit or have discontinued study. If a CSR is written based on data of the second interim analysis the cut-off date of this interim analysis will be used. 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 the Cycle 7 Day 1 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.) is the start of study treatment.

The reference start date for all other, non-safety assessments (i.e., GvHD assessment, survival time, disease progression, response, ECOG performance status, and patient reported outcomes (PRO)) 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] is defined as the last assessment prior to or on the treatment start date.

Baseline for patient reported outcomes is defined as the first assessment prior to or on the study 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 other endpoints is defined as the last assessment or procedure conducted prior to or on the randomization date/day1.

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

On-treatment assessment/event and observation periods

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

- Pre-treatment period: from day of patient's informed consent to the day before first dose of study any study treatment
- On-randomized treatment period: from day of first dose of any study treatment to 30 days after date of last actual administration of randomized treatment (including start and stop date) or end of the randomized treatment period per end of randomized treatment disposition eCRF, whichever is later. For those patients who cross over from BAT to ruxolitinib, the period is from day of first dose of randomized study medication to earlier of (i) 30 days after last dose of randomized study medication, (ii) the day before first dose of cross over treatment. Up to the Cycle 7 Day 1 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.

- On-cross over treatment period: from day of first dose of cross over study medication to 30 days after last dose of cross over study medication) or end of the cross-over treatment period per end of cross-over disposition eCRF, whichever is later
- Post-treatment period: starting at day 31 after last dose of study treatment or the day after end of study treatment per end of treatment disposition eCRFs, whichever is later

Safety summaries (tables, figures) include only data from the on-treatment periods with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). In addition, a separate summary for death including on treatment and post treatment deaths will be provided.

The main comparative safety analyses will use data from the on-randomized treatment period and will be performed for the time period from the day of the first dose up to Cycle 7 Day 1 to avoid potential bias due to different exposure durations related to possible cross over from BAT to ruxolitinib after Cycle 7 Day 1. Furthermore, separate safety summaries will be generated for the entire on-randomized treatment period and the on-cross over treatment period.

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 earlier 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 cGvHD assessment, [redacted] PROs and safety (Table 2-1) by visit. The end of treatment assessment will be mapped into the time points as needed.

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

Time Window	Planned Visit Timing	Time Window Definition
On treatment		
Baseline[1]	On or before Study Day 1	≤ Study Day 1
Cycle 1 Day 8	Study Day 8	Study Days 5 – 11
Cycle 1 Day 15	Study Day 15	Study Days 12 – 18
Cycle 1 Day 22	Study Day 22	Study Days 19 – 25
Cycle 2 Day 1	Study Day 29	Study Days 26 – 39
Cycle 3 Day 1	Study Day 57	Study Days 47 – 67
Cycle 4 Day 1	Study Day 85	Study Days 75 – 95
Cycle 5 Day 1	Study Day 113	Study Days 103 – 123

Cycle 6 Day 1	Study Day 141	Study Days 131 – 151
Cycle 7 Day 1	Study Day 169	Study Days 159 – 179
Cycle 9 Day 1	Study Day 225	Study Days 211 – 239
Cycle 12 Day 1	Study Day 309	Study Days 295 – 323
every 12 weeks	+ 84 day	+84 days to lower and upper bound
Cycle 39 Day 1	Study Day 1065	Study Days 1051 – 1079
Safety follow-up	30 days after last dose	Last dose date + 30

[1] baseline assessment valid if performed prior to or on start date of randomized treatment;
Study Day 1 = reference start date (either randomization date, or start 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 as needed.

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 contact date/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 new cGvHD treatment since discontinuation from study treatment	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
Date of discontinuation from end of treatment pages	No condition.
- cGvHD assessment date - any specific efficacy assessment date if available (e.g. graft failure assessment, hematologic disease relapse/progression assessment)	Non-missing assessment .
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 is 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 (PPS)

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 (CSP).

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

- not corticosteroid refractory cGvHD (i.e. criteria not verified via documented prior medication)
- more than one prior systemic therapy for the treatment of cGvHD, in addition to corticosteroids and CNI
- not moderate or severe cGvHD at randomization
- having transitioned from active aGvHD to cGvHD without tapering off corticosteroids +/− CNI (potential overlap syndrom)
- taking any prohibited medication as specified in this protocol after start of study treatment and before the end of study treatment
- study treatment received is different from treatment assigned by randomization, or no study treatment received at all
- study treatment dispensing errors and/or non-compliance (if considered relevant and documented as protocol deviation)
- cGVHD disease assessment on at least one organ is missing at baseline
- missed Cycle 6 and/or 7 Day 1 cGvHD assessment due to the Covid-19 pandemic

Safety

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.

Pharmacokinetic analysis set (PAS)

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

- take a dose of ruxolitinib
- 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
- for pre-dose samples, have the sample collected before the next dose administration
- The PAS will be used for NCA analysis for patients where extensive PK sampling is obtained and for population PK where sparse sampling is obtained. This analysis set will also be used for any exposure-response analysis [redacted]

Cross-over analysis set

The Cross-over Analysis Set (CAS) comprises all patients randomized to and who receive BAT, who then cross over and receive at least one dose of ruxolitinib. This analysis set will be used for all analyses for cross over patients.

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	INCL02	Not applicable
Safety Set	INCL02	No dose of study treatment
Per-Protocol Set	INCL02, INCL05, INCL07,EXCL01, EXCL02, TRT03, TRT04, TRT05, TRT06, TRT07, TRT08, COMD01, COMD02 and OTH01 and OTH07	No dose of study treatment
Crossover Analysis Set	INCL02	No dose of ruxolitinib
PK Analysis Set	INCL02	No dose of ruxolitinib, No evaluable PK concentration

Note: Based on CINC424D2301_SSD_Edit_checks_and Protocol Deviation Specifications Version 19
INCL02 - 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), US, Asia excluding Japan, Japan
- Chronic GvHD severity (moderate vs. severe)
- Source of grafts
- Criteria for SR-cGvHD (a) a lack of response or disease progression after administration of minimum prednisone 1 mg/kg/day for at least 1 week, b) disease persistence without improvement despite continued treatment with prednisone at >0.5 mg/kg/day or 1 mg/kg/every other day for at least 4 weeks OR c) increase to prednisolone dose to >0.25 mg/kg/day after two unsuccessful attempts to taper the dose)
- Prior cGvHD therapy (corticosteroid only vs. corticosteroid ± CNI)

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). 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 addition to the analyses of the primary efficacy endpoint and the safety summaries displayed for the age group < 18 years, PK will be analyzed separately for adolescents.

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.

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 surface area, body mass index) will be summarized by descriptive statistics (N, mean, median, standard deviation, minimum and maximum).

Baseline stratification factors

The number (%) of patients in each stratum (moderate vs. severe cGvHD) 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 extent of disease in underlying primary disease, stem cell transplant and chronic GvHD.

For underlying primary disease, the analysis will include the following: diagnosis category and subcategory, details of 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 (HLA matched related donor, unrelated UCB (umbilical cord blood), related haploidentical, HLA matched/mismatched URD (unrelated donor)), CMV status, T-cell depleted (Y/N), total nucleated cell dose.

For cGvHD disease history, the analysis will include the following: prior diagnosis of aGvHD, time since initial diagnosis of cGvHD, overall severity of initial cGvHD, steroid refractory cGvHD criteria met (lack of response or disease progression after administration of minimum prednisone 1 mg/kg/day for at least 1 week, disease persistence without improvement despite continued treatment with prednisone at >0.5 mg/kg/day or 1 mg/kg/every other day for at least 4 weeks, Increase to prednisolone dose to >0.25 mg/kg/day after two unsuccessful attempts to taper the dose), cGvHD severity at baseline (study entry), time since steroid refractory cGvHD, cGvHD 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-cGvHD prior to randomization will be summarized by treatment arm.

All data collected at baseline including childbearing 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 who were randomized but not treated, primary reason for not being treated and the number (%) of patients who were treated will be presented. 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 reportable 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 2.2 Analysis sets) 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, 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. Duration of exposure to each BAT regimen (extracorporeal photopheresis (ECP), low-dose methotrexate (MTX). mycophenolate mofetil (MMF), mTOR inhibitors (everolimus or sirolimus), infliximab, rituximab, pentostatin, or imatinib) 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 only, since patients randomized to Investigator's choice of BAT will receive various different categories of therapy.

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

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

The Safety Set and Crossover Analysis Set will be used for all safety summaries.

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	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.

	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.	
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 2.1.1 General definitions

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 Cycle 7 Day 1 and last date of exposure to study treatment (randomized or crossover) will be summarized. Graphical display will also be generated.

Dose reductions, interruptions or permanent discontinuations

The number of patients who have dose reductions, or interruptions, and the reasons, will be summarized for ruxolitinib (randomized and crossover treatment separately). Separate summaries will be generated for patients with dose change due to dose tapering (change of dose due to efficacy of treatment) and patients with dose change due to other reasons (potential safety concerns).The number of patients who have drug permanently discontinued and the reasons, will be summarized by treatment group.

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

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

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

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

Prior prophylaxis includes all systemic treatments that started prior to the diagnosis of cGvHD and administered as prophylaxis for GvHD. The number and percentage of patients who received any prophylaxis (e.g. CNI) prior to randomization will be summarized by lowest ATC class, preferred term and treatment arm using FAS.

Listings will be generated for prophylaxis.

2.4.2.2 Systemic corticosteroid

As per protocol the steroid doses of methylprednisolone will be converted to prednisone equivalents by multiplying the methylprednisolone dose by 1.25. Prednisone doses for each subject are converted to mg/kg/day.

The duration of exposure to systemic corticosteroids 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 the Cycle 7 Day 1

visit and end of on-treatment period. Graphical display will also be generated. These analyses will be based on Safety Set.

2.4.2.3 Calcineurin inhibitors (CNIs) during study treatment

The duration of exposure will be summarized for systemic 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.

2.4.2.4 Additional systemic cGvHD therapy

New additional systemic cGvHD 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.

As per agreement with the Steering Committee members during the SSC meeting on 13th of Nov 2018, initiation of systemic CNI at the time of or after start of study treatment despite absence of lack of efficacy will be counted as new additional systemic treatment. Note: Although the protocol does not explicitly prohibit initiation of systemic CNI at time of randomization or initiation of study treatment potential impact of CNI on efficacy outcome in responders cannot be excluded.

2.4.2.5 Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments, supportive care 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. hemopoietc cytokines and 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 Cycle 7 Day 1 between the ruxolitinib arm and BAT arm in steroid refractory cGvHD patients.

2.5.1 Primary endpoint

ORR at Cycle 7 Day 1, defined as the proportion of patients with complete response (CR) or partial response (PR), according to the NIH Consensus Criteria (Lee 2015). ORR will be calculated based on the FAS using local investigators' overall response assessed at the Cycle 7 Day 1 visit and taking into account initiation or addition of new systemic therapy before this time point.

- **Complete response** is defined as complete resolution of all signs and symptoms of cGvHD in all evaluable organs without initiation or addition of new systemic therapy.
- **Partial response** is defined as an improvement in at least one organ (e.g. improvement of 1 or more points on a 4 to 7 point scale, or an improvement of 2 or more points on a 10 to 12 point scale) without progression in other organs or sites, initiation or addition of new systemic therapies.
- **Lack of response** is defined as unchanged, mixed response, or progression.

A flare of cGvHD is defined as any increase in symptoms or therapy for cGvHD 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 cGvHD flare failure. Patients who fail corticosteroid taper fulfilling either one of the following criteria should initiate additional systemic therapy:

- Re-escalation of 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 <1 mg/kg/day (or equivalent prednisone dose <1.25 mg/kg/day) for a minimum 7 days.

cGvHD Recurrence is defined as the return of cGvHD disease after tapering off study treatment due to response. Following completion of a taper of systemic therapy, if worsening of cGvHD symptoms occur, the patient is allowed to resume treatment for cGvHD as per local institutional practice. For the statistical analyses re-start of treatment for cGvHD is handled in the same way as addition or initiation of new systemic treatment.

A patient will not be considered a responder at Cycle 7 Day 1 if any of the following events occurs prior to the Cycle 7 Day 1 visit:

- Missing overall cGvHD response assessment at Cycle 7 Day 1
- No CR or PR at Cycle 7 Day 1
- Addition of or start of new systemic therapy for cGvHD

2.5.2 Statistical hypothesis, model, and method of analysis

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 Cycle 7 Day 1 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., cGvHD moderate vs severe), will be used to compare ORR between the two treatment groups, at the one-sided 2.5% level of significance.

With protocol amendment 1 an efficacy and safety interim analysis was added and is targeted when 194 (60% of the targeted 324 patients) have completed the Cycle 7 Day 1 visit or discontinued from the study earlier and data of assessments are available. Following the group-sequential methodology defined in amendment 1 the efficacy stopping bound and the respective alpha to be spent will be calculated based on actual information fraction (number of patients included in FAS at the interim analysis divided by 324) using the pre-specified alpha spending function.

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.

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 overall cGvHD response assessments at baseline and Cycle 7 Day 1. The time window for the Cycle 7 Day 1 visit is defined in Table 2-1 above.

No data imputation will be applied. Patients who discontinue study treatment should return for the regular assessments indicated in Protocol Section 7.1. Addition or initiation of a new systemic therapy before Cycle 7 Day 1 in any arm will be considered a treatment failure, and patients will be counted as non-responder in the primary analysis.

2.5.4 Supportive analyses

Supportive analysis will include:

- A detailed description of response rates (CR, PR. Unchanged, mixed response and progression) at Cycle 7 Day 1 by treatment group
- ORR at Cycle 7 Day 1 evaluated with the same analysis conventions as for the primary efficacy analysis using all patients in the PPS (if the PPS differ from the FAS).

- A detailed description of the organ specific response for all organs at Cycle 7 Day 1.

Subgroup analyses for the primary endpoint

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. The subgroups are described in section 2.1.1.
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 Cycle 7 Day 1

ORR at Crossover (CO) Cycle 7 Day 1 is defined as the proportion of crossover patients with complete response (CR) or partial response (PR) at Crossover Cycle 7 Day 1. ORR will be summarized using descriptive statistics (N, %) along with two-sided exact binomial 95% CIs [Clopper and Pearson 1934] using local investigators' overall response assessed at the CO Cycle 7 Day 1 visit and taking into account initiation or addition of new systemic therapy before this time point. Note that response is relative to the last assessment of cGvHD prior to or at the start date of crossover treatment (ruxolitinib).

A patient will not be considered a responder at CO Cycle 7 Day 1 if any of the following events occurs prior to CO Cycle 7 Day 1:

- Missing overall cGvHD response assessment at crossover or CO Cycle 7 Day 1
- No CR or PR at CO Cycle 7 Day 1
- Addition of or start of new systemic therapy for cGvHD

2.6 Analysis of the key secondary objective

The regulatory recommendations for demonstrating additional benefit to patients with SR-GvHD differ between US and ROW: whereas the CHMP/PMDA agreed that Failure free survival (FFS) could be a meaningful measure of clinical benefit, the FDA recommended improvement in patient reported outcomes. Therefore, this study will have two key secondary objectives:

Key secondary objectives

- To compare FFS between ruxolitinib and BAT. FFS is a composite time to event endpoint incorporating the following FFS events: i) relapse or recurrence of underlying disease or death due of underlying disease, ii.) non-relapse mortality, or iii.) addition or initiation of another systemic therapy for cGvHD.

- PRO comparison based on the modified Lee symptom score.

A fixed sequence hierarchical testing strategy will be applied for the primary and the two key secondary endpoints. All tests will be one-sided with significance level alpha=2.5%.

In line with the CHMP/PMDA interactions, FFS is used as the first key secondary endpoint for all regions except the US. The figure below illustrates the sequence that will be used for all regions outside the US. A different testing sequence will be used for the US (FDA), where PRO will be tested before testing FFS.

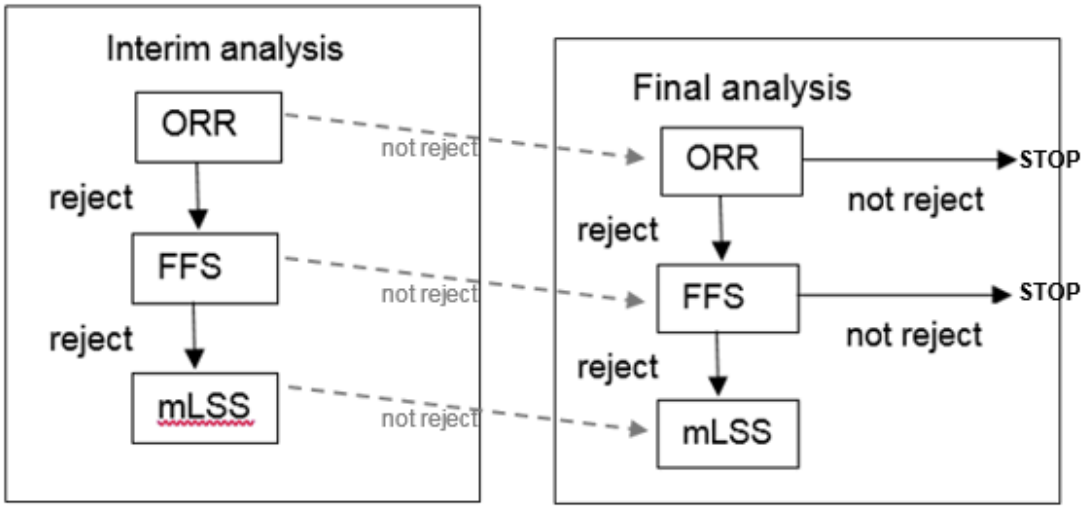
The analysis will be based on the FAS.

The rules how to test key secondary endpoints at the IA were further specified in the DMC Statistical Analysis Plan Amendment 1, an overall hierarchical testing procedure (Figure 2-1) was implemented to control the familywise error rate (FWER) at the pre-specified overall significance level (alpha=0.025) for the primary and the two key secondary endpoints (Glimm et al., Statistics in Medicine, 2010).

Unblinded IA results based on cut-off date 9-Jul-2019 were provided to the DMC in Oct 2019 by the CRO Axio. The efficacy comparison at the IA was based on 196 patients corresponding to an information fraction of 196/324 = 0.605 or 60.5% and a significance level (as per pre-specified alpha spending function) of 0.011763 for the primary and both key secondary endpoints. To avoid any impact on the further conduct of the study the DMC was requested to inform Novartis only in case that 'all three endpoints are positive' when providing their recommendations. The DMC recommended to continue the study as planned.

All randomized patients are included in the primary CSR analysis. Following the overall hierarchical testing procedure, the hypotheses that were not rejected at the IA are to be tested again in this this CSR analyses spending the remaining alpha, i.e. IA results have to be taken into account when interpreting CSR results and are therefore included in this FIR. Assuming that N=329 patients will be included in the CSR analsis the significance level for the tests is alpha=0.01858. (Note: Sum of required significance levels at the IA and the CSR analysis > 0.025 because the repeated test for the CSR is correlated to the test at the IA)

Figure 2-1 Overall hierarchical testing strategy used for all regions except for US



mLSS: responder analysis based on the total symptom score (TSS)

2.6.1 Key secondary endpoints

FFS is defined as the time from date of randomization to the earliest of i) relapse or recurrence of underlying disease or death due to underlying disease, ii.) non-relapse mortality, or iii.) addition or initiation of another systemic therapy for cGvHD. If a patient did not experience any of these events, FFS will be censored at the latest contact data (on or before the cut-off date).

The second key secondary objective (all regions except for US) is to assess the improvement of symptoms based on the total symptom score (TSS) using the modified Lee Symptom Scale. A conceptual overview and scoring rules are given in Section 5.4.3.

A responder is defined as having achieved a clinically relevant reduction from baseline of the total symptom score (TSS). The proportion of responders at the Cycle 7 Day 1 visit (as per visit window defined in Table 2-1), will be compared between treatment groups

A patient will not be considered a responder at Cycle 7 Day 1 if any of the following events occurs prior to the Cycle 7 Day 1 visit:

- Missing or insufficient data to calculate TSS at baseline or Cycle 7 Day 1
- No clinical relevant reduction of TSS at Cycle 7 Day 1
- Addition of or start of new systemic therapy for cGvHD

Multiple data imputation for missing pot-baseline TSS data may be considered to allow additional sensitivity analysis.

2.6.2 Statistical hypothesis, model, and method of analysis

Failure free survival (FFS)

Assuming proportional hazards for FFS, the following statistical hypotheses will be tested:

$$H_{02}: \theta_2 \geq 1 \quad \text{vs.} \quad H_{A2}: \theta_2 < 1$$

where θ_2 is the FFS hazard ratio (ruxolitinib arm versus BAT arm). The analysis to test these hypotheses will consist of a stratified log-rank test at an overall one-sided 2.5% level of significance. The stratification will be based on the randomization stratification factors (i.e. cGvHD moderate vs severe).

Modified Lee symptoms score

Using the total symptom score each patient will be classified as either a responder or a non-responder. A responder definition of 6-7 points has been suggested for the original LSS total score (Lee 2002). For the main analysis in this study a reduction of 7 or more points of the TSS of the mLSS is considered as response. .

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

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

where RR_{rux} and RR_{BAT} are the response rates at Cycle 7 Day 1 in the ruxolitinib and BAT groups, respectively. The Cochrane-Mantel-Haenszel chi-square test, stratified by the randomization stratification factor (i.e., cGvHD moderate vs severe), will be used to compare RR between the two treatment groups, at the one-sided 2.5% level of significance.

Furthermore, the rate of TSS responders and its 95% confidence interval will be presented by treatment group.

Supportive Analyses

The FFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 3, 6, 12, 18 and 24 month FFS estimates and 95% confidence intervals (Brookmeyer and Crowley 1982) will 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.

The cumulative incidence curve of each of the three FFS components (considering the other two components as a competing risk) as well as estimates at 3, 6, 12, 18 and 24 months will also be presented with 95% confidence intervals. More details on these analyses are given in section 5.4.2.

2.6.3 Handling of missing values/censoring/discontinuations

No data imputation will be performed for key secondary endpoints.

2.7 Analysis of secondary efficacy objective(s)

Best overall response (BOR)

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 Cycle 7 day 1 and before the start of additional systemic therapy for cGvHD. ORR will be calculated based on the FAS using local investigators' overall response assessessments.

ORR at Cycle 4 Day 1 (end of Cycle 3)

ORR at Cycle 4 Day 1, defined as the proportion of patients with complete response (CR) or partial response (PR), according to the NIH Consensus Criteria (Lee 2015). ORR will be calculated based on the FAS using local investigators' overall response assessed at the Cycle 4 Day 1 visit and taking into account initiation or addition of new systemic therapy before this time point.

Duration of Response

Duration of response (DOR) is assessed for responders (i.e. all subjects with BOR=CR or PR) only. DOR is defined as the time from first response until cGvHD progression, death, or the date of additional new systemic therapies for cGvHD. Patients without event will be censored at the last contact 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).

Non Relapse Mortality (NRM)

Non-relapse mortality (NRM), defined as the time from date of randomization to date of death not preceded by underlying disease relapse/recurrence. Underlying disease relapse/recurrence is considered as competing events. If a patient is not known to have died or experienced the competing event, then NRM will be censored at the latest date the patient was known to be event-free (on or before the cut-off date).

Incidence of Malignancy Relapse/Recurrence (MR)

Incidence of Malignancy Relapse/Recurrence (MR), defined as the time from date of randomization to date of relapse/recurrence of the underlying disease. Death not proceeded by hematologic relapse/recurrence is considered as competing events. If a patient is not known to have relapse/recurrence or died without relapse/recurrence, then incidence of MR will be censored at the latest date the patient was known to be event-free (on or before the cut-off date).

Reduction of daily corticosteroid dose and successful tapering of all corticosteroids

This will include the assessment of

- Systemic steroid treatment per time interval up to Cycle 7 Day 1
- proportion of patients successfully tapered off all corticosteroids at 6 months

2.7.1 Statistical hypothesis, model, and method of analysis

Best overall response (BOR)

BOR and its 95% confidence interval will be presented by treatment group. P-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will also be presented.

ORR at Cycle 4 Day 1 (end of Cycle 3)

ORR at Cycle 4 Day 1 and its 95% confidence interval will be presented by treatment group. P-value, odds ratio and 95% Wald confidence limits calculated from stratified Cochran-Mantel-Haenszel test will also be presented.

Duration of Response

Duration of response will be calculated for all patients with BOR = CR or PR. Kaplan-Meier curves, medians, 6, 12, 18 and 24 months survival probabilities with 95% confidence intervals will be presented.

Overall survival (OS)

OS will be analyzed according to the randomized treatment group and strata assigned at randomization (cGvHD moderate vs severe). The OS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians, 6, 12 and 24 month survival probabilities 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.

Non Relapse Mortality (NRM)

Cumulative incidence of NRM and derived probabilities at Months 1, 2, 6, 12, 18 & 24 will be estimated, considering underlying disease relapse/recurrence as competing events.

Incidence of Malignancy Relapse/Recurrence (MR)

The cumulative incidence curve for MR and estimates at 3, 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/recurrence and its 95% confidence interval at 3, 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.

2.7.2 Handling of missing values/censoring/discontinuations

Patients with missing overall response assessments that prevent the evaluation of the ORR at Cycle 4 Day 1 and BOR will be considered non-responders on that treatment arm. This includes missing cGvHD assessments at Cycle 4 Day 1. The time window for the Cycle 4 Day 1 visit is defined in Table 2-1 above.

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 the randomized treatment period, 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.

The main comparative safety analyses will use data from the on-randomized treatment period and will be performed for the time period from the day of the first dose up to Cycle 7 Day 1 (= Day 179, the upper bound of the Cycle 7 Day 1 visit window defined in Table 2-1) to avoid potential bias due to different exposure durations and/or related to possible cross over from BAT to ruxolitinib after Cycle 7 Day 1. Furthermore, separate safety summaries will be generated for the entire on-randomized treatment period and the on-cross over treatment period.

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) and post-treatment periods as defined in Section 2.1.1. The 6-month period for the main safety comparison will be derived from the on-randomized treatment period as defined above.

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

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

For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables of on-treatment adverse events which are not serious adverse events with an incidence greater than 5% and on-treatment serious adverse events and SAEs suspected to be related to study treatment will be provided by system organ class and preferred term on the safety set population.

If for the same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE
- more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.8.1.1 AEs adjusted for subject exposure time

In order to account for possible differences in exposure of the ruxolitinib arm relative to the BAT arm due to crossover from BAT to ruxolitinib after Cycle 7 Day 1, 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 the compound ruxolitinib. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, CMQ) may also be used. A CMQ 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 specified AESI, the number and percentage of patients with at least one event of the AESI occurring during the on-treatment period will be summarized. Summaries of these 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.

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.

For laboratory tests where grades are not defined by CTCAE v4.03, results will be categorized as low/normal/high based on laboratory normal ranges.

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:

- Listing of all CTC grade 3 or 4 laboratory toxicities

The results from viral load testing will be summarized by visit and treatment.

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 surface area (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 the following table.

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.

Graphical presentations of vital signs (e.g., body weight, body mass index, systolic blood pressure, diastolic blood pressure) over time for change from baseline may be produced via boxplots based on time windows and corresponding tables displaying the statistics used for the box plots by the selected time points.

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.

2.8.4.3 Pediatric data

Adolescent data will be tabulated and listed, notable values will be flagged.

2.8.4.4 Time to infection

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

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 considering deaths or end of treatment phase (randomized or crossover) without prior infection as competing risks.

2.9 Pharmacokinetic endpoints

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

Plasma samples for PK will be taken at Day 1 (start of treatment), at Day 15 (Cycle 1), to characterize the PK after first dose, and at steady state by non-compartmental analysis. Additional PK samples will be taken at later visits to characterize exposure-efficacy, exposure-safety and population PK modeling as data allows. Concentrations will be expressed in mass per volume units.

Pharmacokinetic parameters of ruxolitinib will be calculated using non-compartmental methods using Phoenix WinNonlin (Pharsight, Mountain View, CA) software. Additional PK parameters may be estimated as needed.

PK parameters

Table 2-6 Non-compartmental PK parameters for ruxolitinib

AUClast	The AUC from time zero to the last measurable concentration sampling time (tlast) (mass x time x volume-1)
AUCinf	The AUC from time zero to infinity (mass x time x volume-1)
AUCtau	The AUC calculated to the end of a dosing interval (tau) at steady-state (amount x time x volume-1)
Cmax	The maximum (peak) observed plasma, blood, serum, or other body fluid drug concentration after single dose administration (mass x volume-1)
Tmax	The time to reach maximum (peak) plasma, blood, serum, or other body fluid drug concentration after single dose administration (time)
Lambda_z	Smallest (slowest) disposition (hybrid) rate constant (time-1) may also be used for terminal elimination rate constant (time-1)
T1/2	The elimination half-life associated with the terminal slope (λz) of a semi logarithmic concentration-time curve (time). Use qualifier for other half-lives
CL/F	The total body clearance of drug from the plasma (volume x time-1)
Vz/F	The apparent volume of distribution during terminal phase (associated with λz) (volume)

Descriptive statistics (n, arithmetic mean, CV% mean, standard deviation (SD), median, geometric mean, CV% geo-mean, minimum and maximum) will be presented by treatment 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 by treatment using the Full analysis set.

Population PK approach

The population pharmacokinetic analysis will be not included into the clinical study report and will be provided in a separate report.

Data handling principles

Plasma concentration values below the limit of quantification (BLQ) will be set to zero by the Bioanalyst, and will be displayed as zero in the listings and flagged. BLQ 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 coefficient of variation (CV%).

Any missing PK parameter or concentration will not be imputed.

[redacted]

2.11 Patient-reported outcomes

The modified Lee Symptom Scale, FACT-BMT and the EQ-5D will be used to collect data on the patient's disease related symptoms and health-related quality of life. Responses to the Lee Symptom Scale, FACT-BMT and EQ-5D will be generated in accordance with the respective scoring manual. Scoring rules for the modified Lee symptom scale are also described in Appendix 5.4.2.2. Further details on the scoring of FACT-BMT and EQ-5D are given in Appendix 5.4.2.3 and Appendix 5.4.2.4.

The main comparative analyses for the modified Lee symptom score is described in Section 2.6 Descriptive statistics (mean, standard deviation, median, minimum, and maximum) will be used to summarize the scored scales at each scheduled assessment time point for all instruments. 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.

No imputation will be applied if the total or subscale scores are missing at a visit.

Resource utilization

Data relating to Resource Utilization (described in Protocol Section 7.2.5) will be used for the purpose of economic evaluation which may be analyzed and reported as a separate activity

Data relating to Resource Utilization will be tabulated for the on-treatment period, all data will be listed.

[redacted]

2.14 Interim analysis

One early safety interim analysis and an efficacy and safety analysis based on 60% of the targeted patients will be performed.

An early safety analysis will be generated when safety data on the first 80 randomized patients who have completed Cycle 4 Day 1 are available. No efficacy data will be generated for this safety analysis.

An efficacy and safety analysis will be performed when 194 (60% of the targeted 324 patients) have completed the Cycle 7 Day 1 visit or discontinued from the study earlier and data of assessments are available. Following a 2-look group sequential design a rho-spending function with parameter rho=1.5 will be used as alpha spending function, for which operational characteristics are given in the Table 2-7. If the number of patients at the IA is not exactly 194 at the time of the interim cut-off, the efficacy stopping bound will be recalculated based on the pre-specified alpha spending function.

Table 2-7 Operational characteristics at the interim analysis

With 194 patients (60%) patients and final analysis

	No. of patients (percent)	Alpha spent[#] (cumulative)	Required odds ratio for significance (if design assumption correct)	Corresponding ORR for Rux (if assumed ORR for BAT correct)	Simulated rejection rates under H1 (cumulative)	
					Reject H0	Reject H1
assuming ORR for BAT=0.58 (58%), odds ratio=2.35						
Interim analysis	194 (59.9)	0.012	2.05	73.88%	66.37%	0%
Final analysis	324 (100)	0.025	1.66	69.65%	93.34%	6.66%
assuming ORR for BAT=0.62 (62%), odds ratio=2.35						
Interim analysis	194 (59.9)	0.012	2.10	77.39%	62.94%	0%
Final analysis	324 (100)	0.025	1.69	73.39%	91.64%	8.36%
assuming ORR for BAT=0.66 (66%), odds ratio=2.35						
Interim analysis	194 (59.9)	0.012	2.17	80.79%	59.25%	0%
Final analysis	324 (100)	0.025	1.73	77.05%	89.36%	10.64%

Based on sequential CMH test with one-sided cumulative alpha=0.025 (2.5%), assuming same ORR and 50% of patients in each stratum)

[#]derived from alpha spending function (rho spending function with rho=1.5)

Note: derived from Table 10-2 of the amended study protocol, in protocol Table 10-2 one digit for the values in the column alpha spent was missing (e.g. 0.25 instead of the correct 0.025), typos were corrected in this table.

The interim analyses will be performed by an independent statistician and programmer (not involved with the conduct of the study). Further details will be described in the DMC charter. The results of the interim analyses will be provided to the DMC by the independent statistician.

Early PK analysis

To compare the exposure of 10 mg BID in the SR cGvHD population to the known exposure in MPN patients, the early extensive ruxolitinib PK data on the first 8 patients (adults and 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.

3 Sample size calculation.

3.1 Primary analysis

The sample size calculation is based on the primary variable ORR on Cycle 7 Day 1. In a meta-analysis Olivieri et al. (Lancet Hematology, 2015) obtained an estimated pooled effect

size for ORR for systemic treatment of SR-cGvHD of 0·66 (95% CI 0·62–0·70). However, as most studies in this analysis did not use objective response criteria it can be expected that application of the NIH Consensus to assess ORR on Cycle 7 Day 1 would lead to lower response rates. Sample size calculations were performed to achieve 90% power for different scenarios (Table 3-1), assuming a targeted odds ratio of 2.35 and 2.5, respectively. A sample size of 324 patients is considered as reasonable for this study.

With sample size 162 patients in each treatment arm and assuming that the assumptions made for sample size calculations are correct (e.g. ORR=0.66 on Cycle 7 Day 1 for BAT for each stratum) an observed odds ratio greater than or equal to 1.68 would achieve statistical significance.

Table 3-1 Sample size for different scenarios

BAT ORR	Ruxolitinib ORR	Odds ratio	Total sample size*
0.58	0.775	2.5	N=247
0.62	0.803	2.5	N=265
0.66	0.829	2.5	N=289
0.58	0.764	2.35	N=278
0.62	0.793	2.35	N=298
0.66	0.820	2.35	N=324

*based on stratified one-sided CMH test (alpha=2.5%, assuming same ORR and 50% of patients in each stratum)

3.2 Power for analysis of key secondary variables

The 6-month FFS estimate based on published work is expected to be 0.56 (Inamoto 2013). It is assumed that treatment with ruxolitinib might reduce the risk of FFS with expected hazard ratio = 0.64 (considering a 6 months FFS rate of 0.69 for ruxolitinib and 0.56 for BAT).

To ensure 90% power for a one-sided log rank test with α=2.5% significance level a total numbers of 212 FFS events would be required. Assuming that enrolment will continue for 18 months with increasing enrolment during the first 6 months and at a uniform rate about 20 patients per month thereafter, a total of N=312 patients would needed to be randomized to observe the targeted 212 FFS events at about 25 months after the randomization date of the first patient. Thus the targeted sample size of N=324 patients is considered adequate for the analysis of FFS.

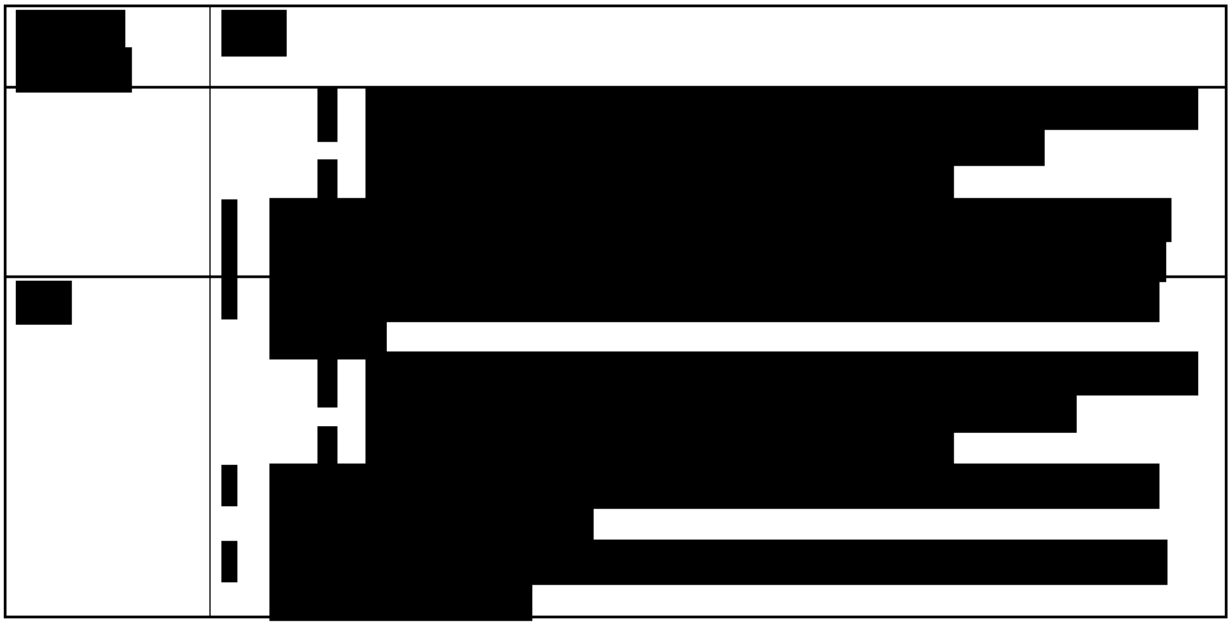
4 Change to protocol specified analyses

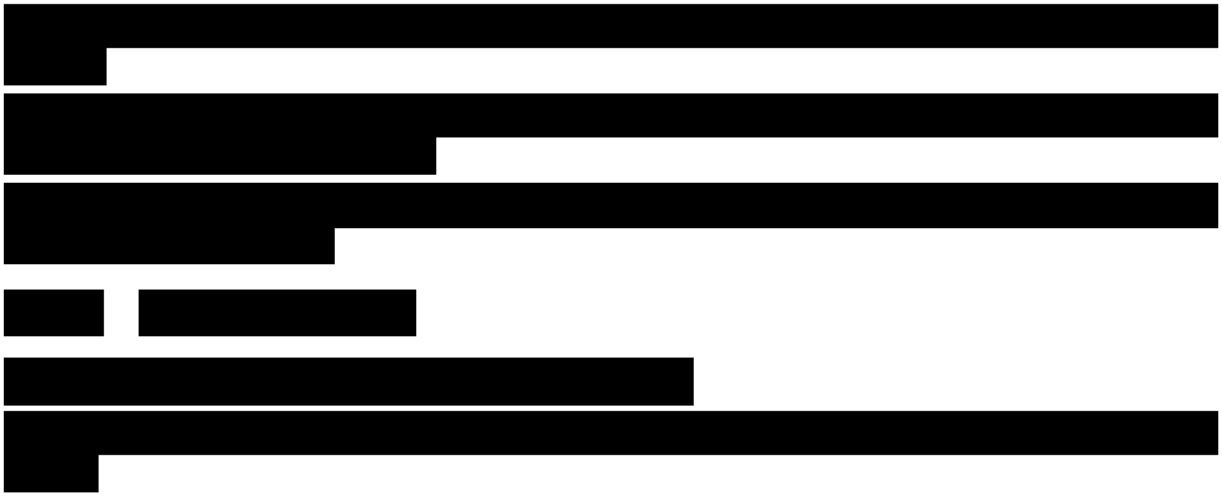
No relevant changes were made between the protocol statistical section and the SAP.

[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]





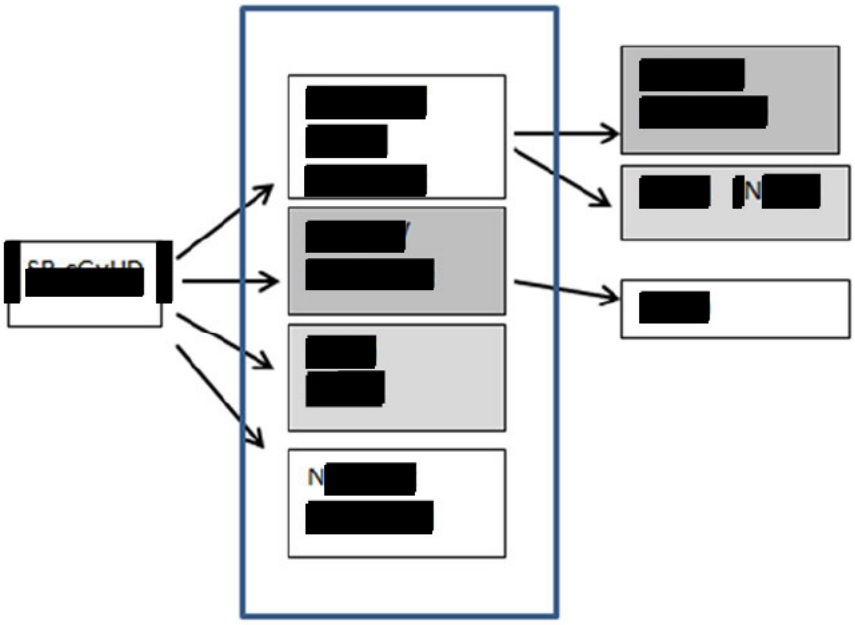
[redacted]

[redacted]

covariates (both categorical and continuous). The odds ratio will be presented with 95% Wald

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

_____ + _____ + _____ =_____=**FACT-BMT TOI**
(**PWB** score) (**FWB** score) (**BMTS** score)

[redacted]

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

[redacted]

[redacted]

[redacted]

6 Reference

1. Armitage P, McPherson CK and Rowe BC (1969). Repeated significance tests on accumulating data. J .R. Statist. Soc. A, 132, 232-244.
2. Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41.
3. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrical, 26, 404-413.
4. Collet D (1994). Modelling survival data in medical research. London, Chapman & Hall.
5. Dmitrienko A., Molenberghs G., Chuang-Stein Ch., Offen W. (2005). Analysis of clinical trials using SAS. SAS Institute Inc., SAS press.
6. Glimm, E., Maurer, W., and Bretz, F. (2010). Hierarchical Testingof Multiple Endpoints in Group Sequential Trials. Statistics in Medicine, 29, 219–228.
7. 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.