NOVARTIS

Clinical Development

INC424/ruxolitinib/Jakavi®

REALISE (CINC424A2411) / NCT02966353

A multicenter phase II, open label, single arm study to evaluate the efficacy and safety of Ruxolitinib in the treatment of anemic myelofibrosis patients.

Statistical Analysis Plan (SAP) Amendment 1

Author:	Trial Statistician, [redacted]

Document type:	SAP Documentation
Document status:	Final
Release date:	12-Mar-2019
Number of pages:	29

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
12 Mar 2019	Before DBL	To modify visit windowing rules	Analysis Visit window table modified to exclude EOT in visit window	Section 2.1.1
		To remove text	Removed text for "Significant non-drug therapies" as it is not captured in the study	Section 2.4.2
		To add Spleen response definition and analysis by IWG-MRT criteria and also to add an analysis for primary endpoint at any time during the study.	Add definition and analysis of spleen response as per IWG MRT Criteria as per the recommendation from steering committee	Section 2.5.4
		To add an analysis on transfusion requirements amd modify some text	Added clarification that unit of blood transfusion will be used for analysis and not the episode of transfusion; An analysis for transfusion dependent rate added for patient who were not TD at baseline	Section 2.7.2
		To add some decription on Adverse Event of Specail Interest (AESI)	Add a breif discription on AESI and a listing is proposed for AESI	Section 2.8.1.1
		To remove analysis	Shift tables based on low/normal/high ranges removed	Section 2.8.3
		To remove analysis	Shift table for baseline ECOG score vs worst post-baseline ECOG score removed. ECOG score over time will only be listed	Section 2.8.4.2
		To add description of PRO related derivations and few analysis	Added analysis on PRO similar to spleen response at Week 24	Section 2.11

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
		To add description of DIPSS score calculation	Added definition and criteria to calculate DIPSS risk scores and risk groups	Section 5

Table of contents

List of abbreviations

AE	Adverse event
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Classification
CRO	Clinical Research Organization
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DI	Dose Intensity
DIPSS	Dynamic International Prognostic Scoring System
ECOG PS	Eastern Cooperative Oncology Group Performance Status
EOT	End of Treatment
FAS	Full Analysis Set
eCRF	Electronic Case Report Form
ICF	Informed Consent Form
IWG-MRT	International Working Group for Myelofibrosis Research and Treatment
LCM	Lower Costal Margin
MedDRA	Medical Dictionary for Drug Regulatory Affairs
MF-7	Myelofibrosis 7 Item Symptom Scale
MFSAF	Myelofibrosis Symptom Assessment Form
NR	Not reached
PD	Pharmacodynamic
NCI	National Cancer Institute
PDI	Planned Dose Intensity
PET-MF	Post-Essential Thrombocythemia Myelofibrosis
PGIC	Patient Global Impression of Change
PK	Pharmacokinetics
PMF	Primary Myelofibrosis
PPV-MF	Post-Polycythemia Vera-Myelofibrosis
PT	Preferred Term
PPS	Per-Protocol Set
PRO	Patient-reported Outcomes
RDI	Relative Dose Intensity
SAE	Serious Adverse Event
TBIL	Total Bilirubin
TD	Transfusion Dependent
SAP	Statistical Analysis Plan
SOC	System Organ Class
TFLs	Tables, Figures, Listings
TD	Transfusion Dependent

TI	Transfusion Independent
ULN	Upper Limit of Normal
WBC	White Blood Cell

1 Introduction

This SAP describes the detailed statistical methodology of the study CINC424A2411: A Phase II study designed to assess the efficacy and safety of ruxolitinib in patients with Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF), or Post-Essential Thrombocythemia Myelofibrosis (PET-MF).

Data will be analyzed by Novartis according to the Data analysis Section 10 of the study protocol which is available in Appendix 16.1.1 of the clinical study report. Important information is given in the following sections and details are provided, as applicable, in Appendix 16.1.9 of the clinical study report.

1.1 Study design

This is a multicenter phase II, open label, single arm study to assess the efficacy and safety of ruxolitinib in patients with Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF), or Post-Essential Thrombocythemia Myelofibrosis (PET-MF) who have splenomegaly that is equal to or greater than 5 cm below the left costal margin and with anemia defined by Hb less than 10 g/dL.

The target number of patients for this trial is 50. The primary anlysis will be performed when all enrolled patients have completed the Week 24 visit or have discontinued from the study prior to Week 24.

No interim analysis is planned for this study.

1.2 Study objectives and endpoints

Objectives and related endpoints are described in Table 1-1 below:

Table 1-1 Objectives and related endpoints

Objective	Endpoint
Primary	
To determine the spleen length response rate at Week 24.	Proportion of patients achieving a 50% reduction in spleen length at Week 24
Secondary	
To evaluate safety.	Safety will be assessed by Frequency and severity of adverse events and serious adverse events and AEs leading to discontinuations. Changes in hematology/biochemistry parameters over time.
To determine the spleen length response rate at Week 48.	Proportion of patients achieving a 50% reduction in spleen length at Week 48.

To evaluate the effect of ruxolitinib on spleen length.	Percent change from baseline in spleen length over time.
To evaluate the effect of ruxolitinib on symptoms.	Summary of the Modified MFSAF v2.0 and MF-7 over time.
To evaluate the effect of ruxolitinib on Patient Global Impression of Change (PGIC).	PGIC at each visit where measured.
To evaluate the effect of ruxolitinib on transfusion requirements.	Summary of transfusions over time. For Transfusion Dependent (TD)* patients, the following will be analyzed: Transfusion independence (TI) rate (requiring no transfusion for ≥12 weeks at any time). Transfusion response rate (not TD: having 5 or less transfusion for ≥12 weeks at any time).

[redacted]

2 Statistical methods

2.1 Data analysis general information

With SAS® ver. 9.4 or higher, the analysis for this study will be conducted by statisticians and statistical programmers from Novartis or designated CRO after all patients have either completed or have been prematurely discontinued from the study.

Unless otherwise specified, qualitative data (e.g., gender) will be described using frequency and percentages, a missing category will be included as applicable; while quantitative data will be described using descriptive statistics: n, mean, standard deviation, median, minimum, and maximum.

Unless otherwise stated, summary tables/figures/listings will be on all patients included in the population under consideration.

All statistical analyses will be performed using all data collected in the database up to data cut-off date. Only 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.

All the events with start dates before or on the cut-off date and end date after the cut-off date will be reported as 'continuing at the cut-off date'. 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 appear as missing in listings.

The final analysis will occur at the end of the study. All available data from all patients will be analyzed and summarized in a final CSR.

2.1.1 General definitions

Study treatment

Study treatment for this study is "ruxolitinib".

Date of enrollment

In this study, the date of enrollment is the date first ICF is signed.

Date of the first/last administration of study treatment

Date of the first administration of study treatment is derived as the first date when a non-zero dose of study treatment is administered.

Date of the last administration of study treatment is derived as the last date when a non-zero dose of study treatment is administered.

Reference start date

Reference start date for this study will be defined as the date of the first administration of study treatment.

Study day

The study day for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption etc.) will be calculated as the difference between the date of the event (onset date of an event, assessment date etc.) and the start of study treatment plus 1. The first day of study treatment is therefore study Day 1. Example: if start of study treatment is on 05-Jan-2014 and start date of an adverse event is on 09-Jan-2014 then the study day of the adverse event onset is 5. For safety assessments before start of study treatment, the study day is negative and derived by (date of event – start date of study treatment). For example: if start of study treatment is on 05-Jan-2014 and date of lab measurement is on 02-Jan-2014 then the study day of the laboratory abnormality is -3. Note, the day of start of study treatment is day 1, and the day before the date of first study treatment is day – 1, not day 0.

Units conversion

If duration is reported in weeks, duration in days will be divided by 7.

Baseline

Baseline is the result of an investigation describing the "true" state of the subject before start of study treatment administration.

The last available assessment on or before the date of start of study treatment (as defined in Section 2.1.3) is taken as "baseline" assessment. In case time of assessment and time of treatment start is captured, the last available assessment before the treatment start date/time is used for baseline.

On-treatment assessment/event and observation periods

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

1. pre-treatment period: from day of subject's first informed consent to the day before first administration of study treatment
2. on-treatment period: from date of first administration of study treatment to 30 days after date of last actual administration of any study treatment (including start and stop date)
3. post-treatment period: starting at day 30+1 after last administration of study treatment.

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

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

Analysis Visit Windows

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 2 assessments within a time window are equidistant from the target date, then the earlier of the 2 assessments will be used. If there are 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.

Analysis Visit windows

Assessment	Target day of assessment	Time Interval
Baseline	1	≤ Day 1
Week 4	28	Day 2 to day 42
Week 8	56	Day 43 to day 70
Week 12	84	Day 71 to day 98
Week 16	112	Day 99 to day 126
Week 20	140	Day 127 to day 154

Week 24	168	Day 155 to day 210
Week 36	252	Day 211 to day 294
Week 36+(12*k)	d=k*84+252, k=1,2,3,etc.	Day d-41 to day d+42

2.2 Analysis sets

2.2.1 Full Analysis Set

The Full Analysis Set (FAS) comprises all patients to whom study treatment (ruxolitinib) has been assigned and who received one dose of study treatment (ruxolitinib).

2.2.2 Safety set

The Safety set is the same as the FAS.

2.2.3 Per-Protocol set

The Per-Protocol Set (PPS) consists of a subset of the patients in the FAS who had no major protocol deviation.

The following major protocol deviations will lead to exclusion from the PPS:

- Patient with less than 12 years of age at screening
- Patients with unconfirmed diagnosis of PMF, PPV-MF or PET-MF, as reviewed by clinical team
- Patient without palpable splenomegaly or with a spleen length below 5 cm at baseline
- Patient with hemoglobin >= than 10 g/dL at baseline
- Patient had prior treatment with any JAK1 or JAK2 inhibitor
- Patient has known hypersensitivity to ruxolitinib or other JAK1/JAK2 inhibitors, or to their excipients
- Treatment with another investigational medication or treated with an investigational medication within 30 days or 5 half-lives (whichever is longer) prior to the first dose of study drug
- Patient unable to comprehend or are unwilling to sign an ICF or Patient did not sign ICF or amended ICF
- Concomitant use of another JAK inhibitor
- Use of investigational medication not approved for any indication within 30 days or 5 half-lives, whichever is longer, prior to the first dose of study drug and until treatment discontinuation
- Use of hydroxyurea, interferon, thalidomide, busulfan, lenalidomide, or anagrelide on Day 1 of ruxolitinib therapy and at any time during participation in the study
- Use of potent CYP3A4 inducers in combination with ruxolitinib

2.2.4 Subgroup of interest

Not Applicable.

2.3 Patient disposition, demographics and other baseline characteristics

Demographic and other baseline data (including disease characteristics) will be listed and summarized descriptively for the FAS.

Categorical data will be presented as frequencies and percentages. For continuous data, mean, standard deviation, median, 25th and 75th percentiles, minimum, and maximum will be presented.

Relevant medical histories and current medical at baseline will be summarized separately for ongoing and historical medical conditions by system organ class and preferred term, for the FAS.

2.3.1 Patient disposition

Enrollment by country and center will be summarized for all screened patients and also by treatment arm using the FAS. The number (%) of treated patients included in the FAS will be presented.

The following summaries will be provided (with % based on the total number of FAS patients):

- Number (%) of patients enrolled,
- Number (%) of patients who completed the study treatment,
- Number (%) of patients who are still on-treatment
- Number (%) of patients who discontinued the study treatment and reasons for discontinuation,
- Number (%) of patients who discontinued the study and reasons for discontinuation.

Disposition data will also be listed.

Protocol deviations

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

Analysis sets

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

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

2.4.1 Study treatment / compliance

Duration of exposure, actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized for Safety Set.

Duration of exposure will be categorized into time intervals; frequency counts and percentages will be presented for the number (%) of patients in each interval. The number (%) of patients who have dose changes or interruption, and the reasons, will be summarized for the Safety Set.

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

Duration of exposure to study treatment:

Duration of exposure to study treatment is considered by taking into account the duration of exposure to the investigational drug:

> Duration of exposure (days) to study treatment = [(date of last non-zero dose administration of study treatment) – (date of first non-zero dose administration of study treatment) + 1 day].

(see Section 2.1.1.3 for definition for the dates first/last administration).

The duration includes the periods of temporary ruxolitinib interruption (s).

The duration of exposure will be summarized by descriptive statistics, as well as by duration category (<= 4 Weeks, >4 – <= 8 Weeks, >8 – <= 12 Weeks, >12 – <= 16 Weeks, >16 – <= 20 Weeks, >20 – <= 24 Weeks, >24 – <= 36 Weeks, and in interval of 12 Weeks thereafter).

Duration of on study:

The duration of on study is defined as the time interval from the enrollment date to the last contact date.

> Duration of on study (days) = [(end of study date – (enrollment date) +1 day]

The duration of study will be summarized by descriptive statistics, as well as by duration category (<= 4 Weeks, >4 – <= 8 Weeks, >8 – <= 12 Weeks, >12 – <= 16 Weeks, >16 – <= 20 Weeks, >20 – <= 24 Weeks, >24 – <= 36 Weeks, and in interval of 12 Weeks thereafter).

Cumulative dose:

Cumulative dose of a study treatment is defined as the total dose given during the study treatment exposure and will be summarized for Safety Set.

The **planned cumulative dose** for a study treatment component refers to the total planned dose as per the protocol up to the last date of investigational drug administration.

The **actual cumulative dose** refers to the total actual dose administered, over the duration for which the subject is on the study treatment as documented in the Dose Administration eCRF.

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

The actual cumulative dose is the sum of the non-zero doses recorded over the dosing period and the planned cumulative dose is the planned starting dose summed over the same dosing period.

Dose intensity and relative dose intensity

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

DI (mg / day) = Actual Cumulative dose (mg) / Duration of exposure to study treatment (day).

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 (day).

Relative dose intensity (RDI) is defined as follows:

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

DI and RDI will be summarized for Safety Set.

Dose reductions, interruptions or permanent discontinuations:

The number of subjects who have dose reductions, permanent discontinuations or interruptions, and the reasons, will be summarized for Safety Set.

'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.

Reduction: A dose change where the prescribed dose level is lower than the previous prescribed dose level or where the actual dose administered/total daily dose is lower than the calculated dose amount based on the prescribed dose. Therefore any dose change to correct a dosing error will not be considered a dose reduction. Only dose change is collected in the CRF, number of reductions will be derived programmatically based on the change and the direction of the change.

2.4.2 Prior, concomitant and post therapies

Prior medications or therapies are defined as drugs or therapies taken and stopped prior to first dose of study treatment. Post-treatment therapies will include all therapies that start after the last non-zero dose of study treatment.

Concomitant medications prior to and after the start of ruxolitinib will be listed and summarized according to the Anatomical Therapeutic Chemical (ATC) classification system by means of frequency counts and percentages. These summaries will include medications starting on or after the start of study treatment or medications starting prior to the start of study treatment and continuing after the start of study treatment.

Any prior concomitant medications or significant non-drug therapies starting and ending prior to the start of study treatment will be listed.

2.5 Analysis of the primary objective

The primary objective of this study is to determine the spleen length response rate at week 24.

2.5.1 Primary endpoint

The primary efficacy variable for this study is the proportion of patients achieving a 50% reduction in spleen length at Week 24. The FAS will be used for the primary analysis.

2.5.2 Statistical hypothesis, model, and method of analysis

The assessment of primary efficacy of study treatment will be based on the calculation of the observed proportion of patients with spleen length response at Week 24 and its posterior distribution using a Beta-binomial model. A patient will be a responder if he/she achieves a 50% reduction in spleen length at Week 24.

Study treatment will be declared efficacious if the following criteria are met:

a. Observed proportion of patients with ≥50% reduction in spleen length at Week 24 ≥25%

b. Probability of true response rate ≤15% (considered not clinically meaningful) is less than 10%

With 50 patients, criteria ‘a’ and ‘b’ will be met if the observed responder rate is ≥25%, i.e., at least 13 patients out of 50 patients have a response at Week 24 which will give a probability of “not being clinically meaningful (response ≤15%)” of 0.0243.

The primary analysis will be performed when all the enrolled patients have been treated with ruxolitinib for 24 weeks or discontinued prior to Week 24. The point estimate of the proportion of patients achieving ≥50% reduction in spleen length at Week 24 along with corresponding exact 95% confidence interval using Clopper and Pearson exact method will be presented.

2.5.3 Handling of missing values/censoring/discontinuations

For the primary analysis, patients with missing baseline spleen length will be excluded from the analysis as change from baseline cannot be calculated.

Patients with missing spleen length at Week 24 or who withdraw earlier from the study will be considered as a non-responder.

2.5.4 Supportive and Sensitivity analyses

The primary analysis (as described in Section 2.5.2) will also be repeated on the PPS. Furthermore, the primary endpoint will be analyzed on patients with non-missing spleen length at both baseline and Week 24. Proportion of patients achieving a 50% reduction in spleen length

at any time during the study will also be reported along with corresponding exact 95% confidence interval using Clopper and Pearson exact method.

IWG-MRT response criteria will also be used to analyze spleen length response rate at Week 24. Response as per IWG-MRT criteria is defined as follows:

- A baseline splenomegaly that is palpable at 5-10 cm, below the LCM at baseline, becomes not palpable, post-baseline or
- A baseline splenomegaly that is palpable at >10 cm at baseline, below the LCM, decreases by ≥50%, post-baseline
- A baseline splenomegaly that is palpable at <5 cm at baseline, below the LCM, is not eligible for spleen response

Patients meeting response criteria as per above definition will be considered as responders or else will be non-responders. The point estimate of the proportion of responders along with corresponding exact 95% confidence interval using Clopper and Pearson exact method will be presented at Week 24. The same analysis will be repeated for proportion of responders as per IWG-MRT response criteria for spleen at any time during the study.

2.6 Analysis of the key secondary objective

There is no key secondary objective for this study.

2.7 Analysis of secondary efficacy objective(s)

All secondary efficacy analyses will be performed on the FAS unless otherwise specified.

2.7.1 Secondary endpoints

The secondary efficacy variables for this study are:

- the proportion of patients achieving a 50% reduction in spleen length at Week 48
- percent change from baseline in spleen length over time
- summary of transfusions over time

2.7.2 Statistical hypothesis, model, and method of analysis

The spleen length response rate at Week 48 will be evaluated. The point estimate of the proportion of patients achieving ≥50% reduction in spleen length at Week 48 along with corresponding exact 95% confidence interval using Clopper and Pearson exact method will be presented.

Percent change from baseline in spleen length over time will be evaluated to assess the effect of ruxolitinib on spleen length. The analysis will be performed for each visit and at end of treatment. This parameter is computed by dividing the difference of spleen length at assessment time and at baseline by the length at baseline multiplied by one hundred.

To evaluate the effect of ruxolitinib on transfusion requirements, the following will be summarized:

- Number and percentage of transfusions over time
- Proportion of patients who are transfusion dependent (TD). Transfusion dependence is defined as per IWG-MRT criteria: 6 or more units of transfusion in 12 weeks prior to baseline.
- For transfusion dependent (TD) patients at baseline, the following will be analyzed:
  - transfusion independence (TI) rate (requiring no unit of transfusion for ≥12 weeks at any time during the study)
  - transfusion response rate (not TD: having 5 or less units of transfusion for ≥12 weeks at any time during the study)
- For patients who were not transfusion dependent (TD) at baseline, the following will be analyzed:
  - transfusion dependence (TD) rate (requiring 6 or more units of transfusion for ≥12 weeks at any time during the study)

2.8 Safety analyses

For all safety analyses, the Safety set will be used, unless specified otherwise.

2.8.1 Adverse events (AEs)

AE summaries will include all AEs occurring during on-treatment period. All AEs collected in the AE eCRF page will be listed along with the information collected on those AEs e.g. AE relationship to study drug, AE outcome etc. AEs starting during the post-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 following adverse event summaries will be produced by treatment; overview of adverse events and deaths (number and % of subjects who died), with any AE, any SAE, 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 (drug withdrawn), leading to dose interruption/adjustment (dose increased or dose reduced), requiring additional therapy (concomitant medication or non-drug therapy) and leading to fatal outcome. In addition, a summary of serious and non-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) as per EudraCT requirements.

2.8.1.1 Adverse events of special interest / grouping of AEs

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to compound INC424. 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, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad. For each specified AESI, number and percentage of patients with at least one event of the AESI occurring during on treatment period will be summarized.

A Case Retrieval Sheet (CRS; an Excel file) with the exact composition of the adverse events groupings is to be used to map reported adverse events to the AESIs groupings (termed Specific Event Categories (SECs) in the CRS). This file may be updated (i.e. it is a living document) based on review of accumulating trial data. The latest version of the CRS document available at the time of the analyses will be used.

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 for the Safety set, post-treatment deaths will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened subjects.

2.8.3 Laboratory data

The summaries will include all assessments available for the lab parameter collected no later then 30 days after the last study treatment administration date (see Section 2.1.1.3).

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

- Changes from baseline in hematology/biochemistry parameters over time.
- Worst post-baseline CTC grade (regardless of the baseline status). Each subject will be counted only for the worst grade observed post-baseline.
- Shift tables using CTC grades to compare baseline to the worst on-treatment value
- .
- Trends of lab parameter values over time (baseline and selected on-treatment timepoints) should be displayed via boxplots based on time windows and corresponding tables displaying the statistics used for the box plots by the selected time points.

The following listings will be produced for the laboratory data:

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

Liver function parameters

Liver function parameters of interest are total bilirubin (TBIL), ALT, AST and alkaline phosphatase (ALP).

The number (%) of subjects with worst post-baseline values will be summarized:

- ALT or AST > 3xULN
- ALT or AST > 5xULN
- ALT or AST > 8xULN
- ALT or AST > 10xULN
- ALT or AST > 20xULN
- TBL > 2xULN
- TBL > 3xULN
- ALT or AST > 3xULN & TBIL > 2xULN
- ALT or AST > 3xULN & TBIL > 2xULN & ALP < 2xULN (potential Hy's law)

Potential Hy's Law events are defined as those patients with a concurrent occurrence of AST or ALT > 3xULN and TBIL > 2xULN and ALP < 2xULN in the same assessment sample during the on-treatment period. Further medical review has to be conducted to assess potential confounding factors such as, liver metastates, liver function at baseline etc.

2.8.4 Other safety data

2.8.4.1 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: body temperature (°C), systolic and diastolic blood pressure (mmHg).

For analysis of vital signs, notable criteria are provided in Table 1-2 below.

Table 2-1 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
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
Body temperature	>= 39.1	-

The number and percentage of subjects with notable vital sign values (high/low) will be presented for the Safety Set.

A listing of all vital sign assessments will be produced and notable values will be flagged. In the listing, the assessments collected during the post-treatment period will be flagged.

2.8.4.2 ECOG performance status

The ECOG PS scale will be used to assess the physical health of patients, ranging from 0 (most active) to 5 (least active):

Table 2-2 ECOG Performance Scale

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

ECOG scores over time will only be listed.

2.9 Pharmacokinetic endpoints

Not Applicable

2.10 PD and PK/PD analyses

Not Applicable

2.11 Patient-reported outcomes

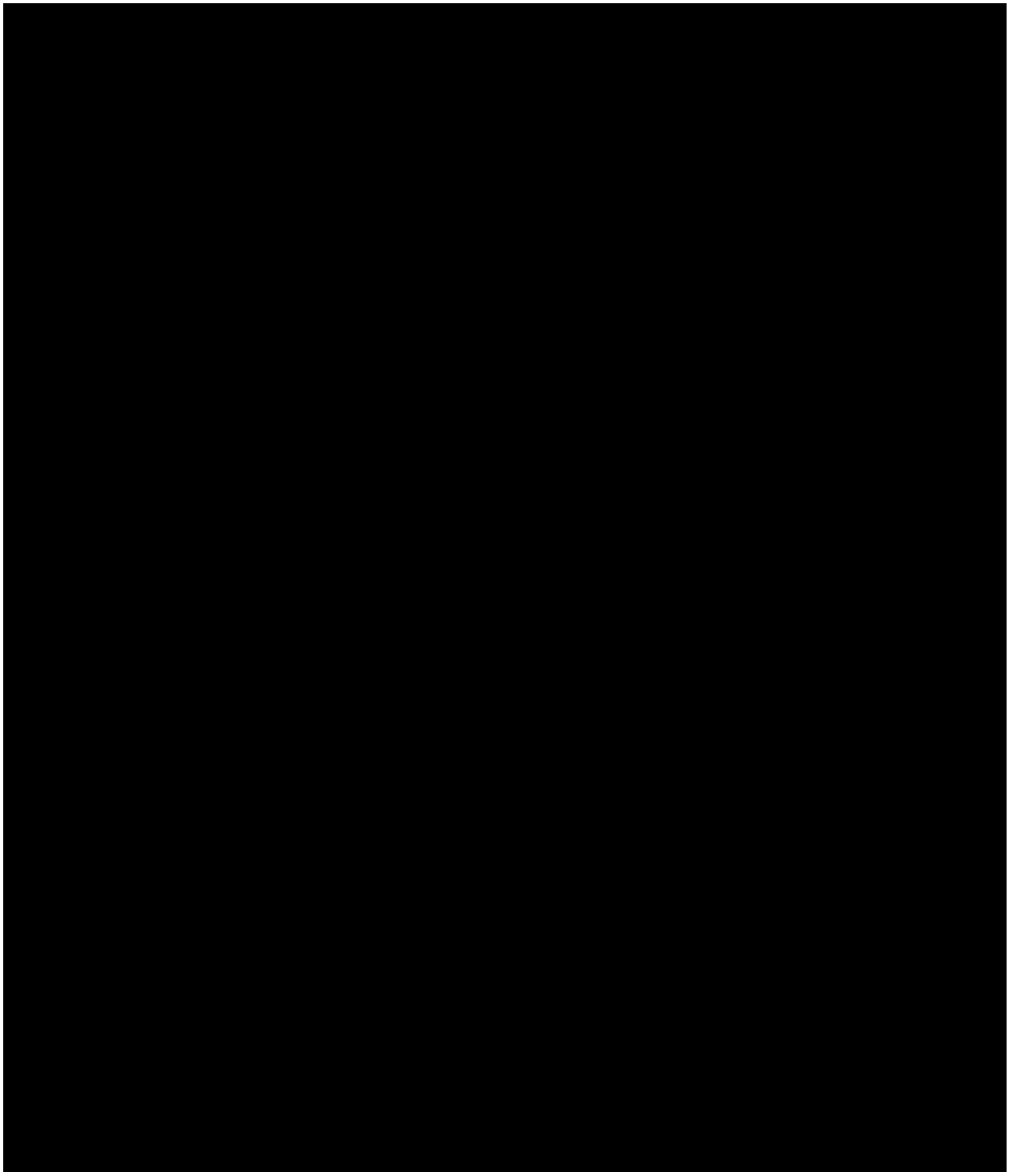
The Myelofibrosis 7 Item Symptom Scale (MF-7) along with Modified MFSAF v2.0 (computed through the addition of a separate question on Inactivity to the data captured with MF7) and the PGIC will be primarily used to collect electronically the patient's disease-related symptoms and patient impression of change. No formal statistical test will be performed.

Missing items data in a scale will be handled based on each instrument manual. No imputation will be applied if the total or subscale scores are missing at a visit. The FAS will be used for analyzing PRO data. Descriptive statistics (e.g., mean, median) will be used to summarize the scored scales at each scheduled assessment time point for the Myelofibrosis 7 Item Symptom Scale (MF-7) and the Modified MFSAF v2.0 score.

Change from baseline in the domain 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. The number and

percentage of patients in each PGIC response category at selected time points will be summarized.

Proportion of patients with at least 50% reduction in MF-7 and modified MFSAF v2.0 from baseline will also be reported at Week 24 along with its 95% CI using exact Clopper-pearson method.



[redacted]

2.13 Other Exploratory analyses

Not Applicable.

2.14 Interim analysis

No formal interim analysis is planned for this trial.

The primary analysis will be performed after all patients have completed Week 24 or discontinued prior to Week 24. Any other interim updates may be provided to fulfill regulatory obligations to comply with post-approval commitments, or for publication purposes.

The final analysis will occur at the end of the study. All available data from all patients will be analyzed and summarized in a final CSR.

3 Sample size calculation

Approximately 50 patients will be enrolled in this study. Success at Week 24 will be declared if the following criteria are met:

a. Observed proportion of patients with ≥50% reduction in spleen length at Week 24 ≥25%

b. Probability of true response rate ≤15% (considered not clinically meaningful) is less than 10%

With 50 patients, criteria 'a' and 'b' will be met if the observed responder rate is ≥ 25%, i.e. if at least 13 patients out of 50 patients have a response at Week 24.

Sample Size	**No. of responders**	**Observed responder rate (%)**	**Probability of true responder rate ≤ 15%**
N=50	11	22	0.1014
	12	24	0.0520
	13	**26**	**0.0243**

	14	28	0.0104
	15	30	0.0041
	16	32	0.0015

The operating characteristics of the design are presented in the table below. This table shows the probability for positive conclusion (i.e. success criteria met at the end of the study) under different true responder rates. When the true responder rate at Week 24 is ≤ 15%, the probability of positive conclusion is ≤ 10% (false success rate). The probability of success is > 90% when the true responder rate at Week 24 is 35%.

Sample Size	**True responder rate (%)**	**Probability of positive conclusion**
N=50	15	0.0301
	20	0.1861
	25	0.4890
	30	0.7771
	35	0.9339
	40	0.9867
	45	0.9982
	50	0.9998

4 Change to protocol specified analyses

Not Applicable.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

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

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

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

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

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

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

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

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

Use Dec31yyyy

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

Use EOT date

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

Use last day of the Month (mm)

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

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

Use the treatment start date

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

5.1.2 AE, ConMeds and safety assessment date imputation.

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

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

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

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

Missing Element	Rule
	• If available month and year < month year of study treatment start date then 15MONYYYY

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

Missing Element	Rule (*=last treatment date plus 30 days not > (death date, cut-off date, withdrawl of consent date))
day, month, and year	Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period *
day	If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

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

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

5.1.2.1 Other imputations

Not applicable

5.2 AEs coding/grading

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

AEs will be assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4.03.

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death. This grading system inherently places a value on the importance of an event, although there is not necessarily proportionality among grades (a grade 2 is not necessarily twice as bad as a grade 1).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. The calculation of laboratory CTC grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTC grades are given in <Novartis internal criteria

for CTC grading of laboratory parameters>. The latest available version of the document based on the underlying CTCAE version 4.03 at the time of analysis will be used.

For laboratory tests where grades are not defined by CTCAE v4.03, results will be graded by the low/normal/high (or other project-specific ranges, if more suitable) classifications based on laboratory normal ranges.

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

Imputation Rules

CTC grading for blood differential is based on an absolute value. However, these data may not be reported as absolute counts but rather as percentage of WBC.

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

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

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

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

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

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

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

5.4 Time windows for box plots

All scheduled/unscheduled assessments should be assigned to time windows. In case of multiple values per window, the one closest to the planned visit date should be used. If 2 values are equidistant to the planned visit date, the selection should be made by selecting the one assessed by central (if any) and otherwise - for multiple central assessments or ECGs / vital signs equidistant to the planned visit - the last value.

Boxplots should display boxes (25^{th} – 75^{th} percentiles) and median (line) as well as joint mean (dot) values. Whiskers extend to the 5^{th} – 95^{th} percentiles with any values outside the whiskers not being displayed. A footnote to the box plot should explain that these values are not displayed. However, all values are used to calculate mean and percentiles displayed on the graph, and

corresponding summary tables should be generated. The number of observations used for this analysis is to be stated at the bottom.

5.5 Dynamic International Prognostic Scoring System (DIPSS)

- The Dynamic International Prognostic Scoring System (DIPSS) can be used to assess a patient's prognosis as conditions change over time.
- DIPSS risk score will be derived for both Primary and secondary MF patients at baseline

Risk Factor	Points		
	0	1	2
Hemoglobin	≥10 g/dL	…	<10 g/dL
WBC count	≤ 25×10^9 L	> 25 ×10^9 L	…
Constitutional symptoms	No	Yes	…
Age	≤ 65 years	> 65 years	…
Peripheral blood blasts	< 1 %	≥ 1 %	…

The risk level is assigned by the total score:

Risk Group	Points	Median Survival, Years
High	5-6	1.5
Intermediate 2	3-4	4
Intermediate 1	1-2	14.2
Low	0	NR

5.6 Statistical models

5.6.1 Primary analysis

Calculation of posterior probability

The efficacy criteria for this study are:

1. Observed proportion of patients with ≥50% reduction in spleen length at Week 24 ≥25%
2. Probability of true response rate ≤15% (considered not clinically meaningful) is less than 10%

Let p denote the proportion of patients with ≥50% reduction in spleen length at Week 24 and follow a beta prior distribution Beta [a,b], where a>0, b>0. Let y out of n patients acheive ≥50% reduction in spleen length at Week 24. Therefore, the posterior distribution of p is Beta(a+ y, b+ n-y) [Spiegelhalter et al. 2004].

A minimally informative unimodal Beta prior [Neuenschwander et al. 2008] Beta [*1/3,1*] will be used. The parameters were chosen so that the posterior probability of the true response rate ≤15% (considered not clinically meaningful) is equal to *0.0243*.

The efficacy criteria will be assessed based on the actual number of patients enrolled in the study. For example, if the total number of patients is 50, the efficacy criteria requires that at least 13 patients who have ≥50% reduction in spleen length at Week 24. In that case, the posterior distribution is Beta (0.333+13, 1+50-13) and the probability of success is > 90% when the true reponder rate is 35%.

Confidence interval for response rate

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

5.6.2 Key secondary analysis

Not Applicable.

6 Reference

1. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrical, 26, 404-413.
2. Neuenschwander B, Branson M and Gsponer T (2008). Critical aspects of the Bayesian approach to Phase I cancer trials. Statistics in Medicine, 27, 2420-2439.
3. Spiegelhalter DJ, Abrams KR and Myles JP (2004). Bayesian Approaches to Clinical Trials and Health-care Evaluation. Chichester, Wiley.